

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2001, 02:43:44 ; Search time 1201.09 Seconds  
(without alignments)  
4327.039 Million cell updates/sec

Title: US-08-224-621-74

Perfect score: 336  
Sequence: 1 CGTCGACATTTTCTATGTC.....CCCCNTTGTTCCCAACCCA 336

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

al number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba1: \*  
2: gb\_ba2: \*  
3: gb\_ba3: \*  
4: gb\_in1: \*  
5: gb\_in2: \*  
6: gb\_in3: \*  
7: gb\_om: \*  
8: gb\_ov: \*  
9: gb\_pat1: \*  
10: gb\_pat2: \*  
11: gb\_ph: \*  
12: gb\_p11: \*  
13: gb\_p12: \*  
14: gb\_p13: \*  
15: gb\_p14: \*  
16: em\_ba1: \*  
17: em\_ba2: \*  
18: em\_fun: \*  
19: em\_htgo\_hum: \*  
20: em\_htgo\_inv: \*  
21: em\_htgo\_rod: \*  
22: em\_htg\_hum1: \*  
23: em\_htg\_hum2: \*  
24: em\_htg\_hum3: \*  
25: em\_htg\_hum4: \*  
26: em\_htg\_hum5: \*  
27: em\_htg\_hum6: \*  
28: em\_htg\_hum7: \*  
29: em\_htg\_hum8: \*  
30: em\_htg\_inv1: \*  
31: em\_htg\_inv2: \*  
32: em\_htg\_other: \*  
33: em\_htg\_rod: \*  
34: em\_hum1: \*  
35: em\_hum2: \*  
36: em\_hum3: \*  
37: em\_hum4: \*  
38: em\_hum5: \*  
39: em\_hum6: \*  
40: em\_hum7: \*  
41: em\_in: \*  
42: em\_om: \*  
43: em\_or: \*

44: em\_ov: \*  
45: em\_pal: \*  
46: em\_ph: \*  
47: em\_p1: \*  
48: em\_ro: \*  
49: em\_sts: \*  
50: em\_sy: \*  
51: em\_un: \*  
52: em\_v1: \*  
53: gb\_sts1: \*  
54: gb\_sts2: \*  
55: gb\_sts3: \*  
56: gb\_sy: \*  
57: gb\_un: \*  
58: gb\_v11: \*  
59: gb\_v12: \*  
60: gb\_htg1: \*  
61: gb\_htg2: \*  
62: gb\_htg3: \*  
63: gb\_htg4: \*  
64: gb\_htg5: \*  
65: gb\_htg6: \*  
66: gb\_htg7: \*  
67: gb\_htg8: \*  
68: gb\_htg9: \*  
69: gb\_htg10: \*  
70: gb\_htg11: \*  
71: gb\_htg12: \*  
72: gb\_htg13: \*  
73: gb\_htg14: \*  
74: gb\_htg15: \*  
75: gb\_htg16: \*  
76: gb\_htg17: \*  
77: gb\_htg18: \*  
78: gb\_htg19: \*  
79: gb\_htg20: \*  
80: gb\_htg21: \*  
81: gb\_htg22: \*  
82: gb\_htg23: \*  
83: gb\_htg24: \*  
84: gb\_htg25: \*  
85: gb\_pr1: \*  
86: gb\_pr2: \*  
87: gb\_pr3: \*  
88: gb\_pr4: \*  
89: gb\_pr5: \*  
90: gb\_pr6: \*  
91: gb\_pr7: \*  
92: gb\_pr8: \*  
93: gb\_pr9: \*  
94: gb\_pr1: \*  
95: gb\_rod: \*  
96: gb\_in4: \*  
97: gb\_pr10: \*  
98: em\_ba3: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	80.6	24.0	119945	92	HSR218J18	AL034370 Human DNA
2	80.6	24.0	162409	67	AC022184	AC022184 Homo sapi
3	78.8	23.5	1846	93	HSNDPG	X65882 H. sapiens m
4	77.2	23.0	1872	93	HSCHRX	X65724 H. sapiens D
5	59.2	17.6	1769	94	MMNDP	X83794 M. musculus
6	59.2	17.6	1774	94	MMNORPTE	X92397 M. musculus
7	41	12.2	63739	77	AC087621	AC087621 Homo sapi
8	41	12.2	115958	85	AC004736	AC004736 Human Chr

c	9	34.8	10.4	171453	61	AC010136	Homo sapi
c	10	3510	9.8	1803016	85	AB033016	Homo sapi
c	11	33	9.8	164036	67	AC018471	Homo sapi
c	12	33	9.8	180359	67	AC022064	Homo sapi
c	13	33	9.8	185161	86	AC006059	Homo sapi
c	14	32.8	9.8	161547	74	AC075330	Homo sapi
c	15	32.8	9.8	177765	74	AC069409	Homo sapi
c	16	32.8	9.8	180606	74	AC073406	Homo sapi
c	17	32.6	9.7	319	92	HS20H4R	Homo sapi
c	18	32.6	9.7	171060	79	AL161728	Homo sapi
c	19	32.6	9.7	180856	79	AL157936	Homo sapi
c	20	32.6	9.7	213741	67	AC022775	Mus muscu
c	21	32.4	9.6	96327	65	AC019700	Drosophila
c	22	32.4	9.6	117927	87	AC013416	Homo sapi
c	23	32.4	9.6	177376	74	AC010120	Drosophila
c	24	32.4	9.6	180490	74	AC069152	Homo sapi
c	25	32.4	9.6	236716	5	AE003754	Drosophila
c	26	32.4	9.6	287870	74	AC073723	Mus muscu
c	27	32.2	9.6	205406	77	AC087900	Mus muscu
c	28	32	9.5	69284	67	AC023185	Homo sapi
c	29	32	9.5	69284	67	AC023185	Homo sapi
c	30	32	9.5	89393	67	AC022667	Homo sapi
c	31	32	9.5	155316	75	AC078902	Homo sapi
c	32	32	9.5	158665	80	AL356138	Homo sapi
c	33	32	9.5	159740	62	AC012063	Homo sapi
c	34	32	9.5	176291	88	AC024610	Homo sapi
c	35	31.8	9.5	39731	86	AC005760	Homo sapi
c	36	31.8	9.5	162896	67	AC023131	Homo sapi
c	37	31.8	9.5	178617	66	AC021529	Homo sapi
c	38	31.8	9.5	182009	72	AC005835	Homo sapi
c	39	31.6	9.4	7681	15	SCYOL155C	X89717 S. cerevisia
c	40	31.6	9.4	15500	14	SCA0GENES	X89717 S. cerevisia
c	41	31.6	9.4	110000	84	LMFICHR36_00	AL499624 Leishmani
c	42	31.6	9.4	132555	87	AC015564	Homo sapi
c	43	31.6	9.4	172423	60	AC008485	Homo sapi
c	44	31.6	9.4	177990	64	AC016385	Homo sapi
c	45	31.6	9.4	184001	81	AL512303	Homo sapi

## ALIGNMENTS

RESULT 1  
LOCUS HSA218J18 119945 bp DNA PRI 23-NOV-1999  
DEFINITION Human DNA sequence from clone 218J18 on chromosome Xp11.3-11.4.  
Contains the NDP (Norrie Disease (Pseudoglioma)) gene and a CCL3  
Splicing Factor pseudogene. Contains ESTs, STSs and GSSs, complete  
sequence.  
VERSION AL034370  
KEYWORDS HTG; CCL3 Splicing Factor; NDP; Norrie Disease; Pseudoglioma.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 119945)  
Wray, P.  
Direct Submission  
Submitted (26-NOV-1998) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
Requests: clonequery@sanger.ac.uk  
This sequence has been finished according to sequence map criteria  
as follows: An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'

feature key.  
IMPORTANT: This sequence is not the entire insert of clone 218J18.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we arrange for a small overlap between  
neighbouring submissions.  
The true right end of clone 27K14 (295125) is at 100 in this  
sequence. This sequence was generated from part of bacterial clone  
contigs of human chromosome X, constructed by the Sanger Centre  
Chromosome X Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/ChrX>  
218J18 is from the library RPI6 constructed at the Roswell Park  
Cancer Institute by the group of Pieter de Jong. For further  
details see <http://bacpac.med.buffalo.edu/VECTOR.pac4>.

## FEATURES

source  
1..119945  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/map="p11.3-11.4"  
/clone="RP6-218J18"  
/clone.lib="RPI-6"  
411..804  
/note="MLT1B repeat: matches 1..394 of consensus"  
1840..1895  
/note="28 copies 2 mer ta 70% conserved"  
2155..7740  
/note="L1PA8 repeat: matches 555..6155 of consensus"  
7888..8034  
/note="FLM1C repeat: matches 2..143 of consensus"  
8908..9255  
/note="MST1 repeat: matches 1..394 of consensus"  
9767..9798  
/note="MST2 repeat: matches 220..251 of consensus"  
9797..9923  
/note="MST3 repeat: matches 391..518 of consensus"  
9929..10110  
/note="MER58A repeat: matches 42..223 of consensus"  
10248..10346  
/note="MLT1G repeat: matches 12..155 of consensus"  
11115..11332  
/note="MIR repeat: matches 32..245 of consensus"  
11868..12249  
/note="R1ggr3(Golem) repeat: matches 1..3035 of  
consensus"  
12459..12548  
/note="MIR repeat: matches 20..130 of consensus"  
12651..13006  
/note="MLT1D repeat: matches 1..314 of consensus"  
13005..13330  
/note="MLT1D repeat: matches 72..358 of consensus"  
13498..13651  
/note="L2 repeat: matches 2447..2592 of consensus"  
14071..14210  
/note="L2 repeat: matches 2605..2750 of consensus"  
14577..14986  
/note="MER74A repeat: matches 1..431 of consensus"  
15386..16018  
/note="L1MB4 repeat: matches 5181..5827 of consensus"  
15972..16516  
/note="L1P repeat: matches 4397..4940 of consensus"  
16519..16720  
/note="AluSg/x repeat: matches 100..295 of consensus"  
16737..16903  
/note="AluSx repeat: matches 35..225 of consensus"  
16904..16998  
/note="L1P repeat: matches 4891..4985 of consensus"  
17144..17247  
/note="L2 repeat: matches 2410..2522 of consensus"  
17665..17883  
/note="L1M4 repeat: matches 4896..5125 of consensus"  
18105..18239  
/note="L1M4 repeat: matches 7694..7824 of consensus"  
18357..18532  
repeat\_region

repeat\_region /note="LIMC4 repeat: matches 7788. .7954 of consensus"  
20579. .20700  
/note="MIR repeat: matches 48. .138 of consensus"  
20701. .20969  
repeat\_region /note="AluJo repeat: matches 9. .281 of consensus"  
20970. .21069  
repeat\_region /note="MIR repeat: matches 138. .247 of consensus"  
22317. .22579  
repeat\_region /note="MER33 repeat: matches 62. .324 of consensus"  
23269. .23575  
repeat\_region /note="AluSg repeat: matches 1. .306 of consensus"  
23834. .24096  
repeat\_region /note="MIR repeat: matches 17. .262 of consensus"  
24827. .24928  
repeat\_region /note="MIR repeat: matches 130. .238 of consensus"  
24961. .25000  
repeat\_region /note="8 copies 5 mer ttgt 85% conserved"  
24961. .24998  
repeat\_region /note="19 copies 2 mer tt 84% conserved"  
25000. .25119  
repeat\_region /note="FLAM\_A repeat: matches 1. .133 of consensus"  
25646. .25750  
repeat\_region /note="L2 repeat: matches 2571. .2683 of consensus"  
25901. .26177  
repeat\_region /note="LTR16A repeat: matches 92. .450 of consensus"  
/complement(join(28998. .30248,38694. .39074,53526. .53726))  
/gene="NDP"  
/note="match: cDNAs X65724 X65882 X92397; match: ESTs  
AA045724 AA670439 W61129 N59262 AA045724 A1129296 W61167  
W6731 R84741 H85821"  
/evidence="not\_experimental"  
/product="NDP (Norrie Disease (Pseudoglioma))"  
/complement(28998. .53726)  
/gene="NDP"  
complement(28998)  
/gene="NDP"  
/complement(29013. .29019)  
/gene="NDP"  
complement(join(30021. .30248,38694. .38867))  
/gene="NDP"  
/note="match: proteins Q00604 P48744"  
/codon\_start=1  
/evidence="not\_experimental"  
/product="NDP (Norrie Disease (Pseudoglioma))"  
/protein\_id="CAA22268.1"  
/db\_xref="GI:3947697"  
/translation="MKRHVLAASFMSILVINGDPTDSKTDSSFINDSPPRCMRHHY  
VDSISHPLYCCKSVLWLAECGHCSDASRSEPLVSFTVLKQPFSSCHCRPPTSK  
LKAFLRRCGSGMLTATYRIILCHCEECNS"  
30396. .30515  
repeat\_region /note="MER5B repeat: matches 58. .178 of consensus"  
30652. .30703  
repeat\_region /note="26 copies 2 mer aa 73% conserved"  
31354. .31662  
repeat\_region /note="AluJo repeat: matches 1. .308 of consensus"  
32205. .32377  
repeat\_region /note="FAM repeat: matches 1. .173 of consensus"  
33007. .33238  
repeat\_region /note="L2 repeat: matches 2177. .2417 of consensus"  
33350. .33381  
repeat\_region /note="16 copies 2 mer ac 100% conserved"  
/complement(33360. .33924)  
/gene="NDP"  
/note="match: GSS AQ042323"  
33571. .33727  
repeat\_region /note="MER5A repeat: matches 9. .189 of consensus"  
33816. .34022  
repeat\_region /note="MIR repeat: matches 7. .213 of consensus"  
34667. .34794  
repeat\_region /note="L2 repeat: matches 2579. .2698 of consensus"  
35507. .35714  
repeat\_region /note="MIR repeat: matches 27. .248 of consensus"  
36253. .36416

repeat\_region /note="MIR repeat: matches 63. .244 of consensus"  
40405. .40533  
repeat\_region /note="MIR repeat: matches 20. .146 of consensus"  
41308. .41535  
repeat\_region /note="L1P4 repeat: matches 5408. .5640 of consensus"  
42540. .43017  
prim\_transcript /note="match: EST AA779268"  
43414. .43506  
repeat\_region /note="L2 repeat: matches 2646. .2741 of consensus"  
44406. .44483  
repeat\_region /note="LIM4B repeat: matches 154. .237 of consensus"  
44932. .45016  
repeat\_region /note="MIR repeat: matches 56. .142 of consensus"  
45355. .45652  
repeat\_region /note="AluJo repeat: matches 1. .294 of consensus"  
48165. .48194  
repeat\_region /note="15 copies 2 mer aa 87% conserved"  
48315. .48514  
repeat\_region /note="match: STS AF020217"  
51372. .51487  
repeat\_region /note="L2 repeat: matches 2589. .2709 of consensus"  
53680. .53741  
repeat\_region /note="31 copies 2 mer ag 90% conserved"  
56812. .57069  
repeat\_region /note="MIR repeat: matches 2. .262 of consensus"  
Query Match 24.0% Score 80.6; DB 92; Length 119945;  
Best Local Similarity 77.7%; Pred. No. 1.6e-15;  
Matches 146; Conservative 0; Mismatches 37; Indels 5; Gaps 4;  
QY 2 GCTGCATCTTTTTCATGCTCTCCCTGCTGCGCTGATGAGATACAGACGAAACG 61  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 38849 GCTGCATCTTTTTCATGCTCTCCCTGCTGCTGATGAGATACAGACGAAACG 38790  
QY 62 GACAGCTGCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 121  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 38789 GACAGCTGCTTCATGATGATGATGATGATGATGATGATGATGATGATGATG 38733  
QY 122 NATCTATACAGCTCACCCTGCTGATGATGATGATGATGATGATGATGATGATG 181  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 38732 GATTCATACG--TCACCATTTGACAGTGTACTCAAGATGACCAAGGCTCTGCT 38675  
QY 182 TCTGTGAC 189  
|||  
Db 38674 GAGGAGAG 38667  
RESULT 2  
AC022184  
LOCUS AC022184 162409 bp DNA HTG 10-SEP-2000  
DEFINITION Homo sapiens chromosome 4 clone RP11-5264 map 4, WORKING DRAFT  
SEQUENCE, 12 unordered pieces.  
ACCESSION AC022184  
VERSION AC022184.3 GI:10047716  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 162409)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bede,F.,  
Boguslavsky,I., Bouckhalter,B., Brown,A., Burkett,G., Castle,A.,  
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Festeror,J.,  
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,  
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Landers,T., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K.,

TITLE  
JOURNAL

## COMMENT

Macdonald, P., Margulis, N., McEwan, P., McGurk, A., McKernan, K.,  
McPheeters, R., Meldrum, J., Menees, L., Morrow, J., Naylor, J.,  
Norman, C.H., O'Connor, T., O'Donnell, P., Oliver, T.M., Peterson, K.,  
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,  
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Rhoman, N.,  
Stojanovic, N., Subramanian, A., Talmas, J., Testaye, S., Theodore, J.,  
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,  
Zimmer, A., and Zody, M.  
Submitted (26-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 10, 2000 this sequence version replaced 91:7107809.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www.seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
Project Information  
Center project name: L2268  
Center clone name: 52\_G\_4

Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 150598 bases at least Q40  
Consensus quality: 156229 bases at least Q30  
Consensus quality: 158583 bases at least Q20  
Insert size: 188000; agarose-*ip*  
Insert size: 161309; sum-of-*contigs*  
Quality coverage: 3.7 in Q20 bases; agarose-*ip*  
Quality coverage: 4.3 in Q20 bases; sum-of-*contigs*

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 12 *contigs*. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the *contigs* are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 6777: *contig* of 6777 bp in length  
6778 6877: gap of 100 bp  
6878 8136: *contig* of 1259 bp in length  
8137 8236: gap of 100 bp  
8237 9996: *contig* of 1760 bp in length  
9997 10096: gap of 100 bp  
10097 12552: *contig* of 2456 bp in length  
12553 12652: gap of 100 bp  
12653 20026: *contig* of 7374 bp in length  
20027 20126: gap of 100 bp  
20127 27970: *contig* of 7844 bp in length  
27971 28070: gap of 100 bp  
28071 36727: *contig* of 8657 bp in length  
36728 36827: gap of 100 bp  
36828 50972: *contig* of 14145 bp in length  
50973 51072: gap of 100 bp  
51073 68989: *contig* of 17917 bp in length  
68990 69089: gap of 100 bp  
69090 91875: *contig* of 22786 bp in length  
91876 91975: gap of 100 bp  
91976 129030: *contig* of 37055 bp in length  
129031 129130: gap of 100 bp  
129131 162409: *contig* of 33279 bp in length.

FEATURES  
SOURCE

1. 162409  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/map="4"  
/clone="RP11-52G4"

misc\_feature  
/clone\_lib="RPCT-11 Human Male BAC"  
1. 6777  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left  
6878. 8136  
/note="assembly\_fragment"  
8237. 9996  
/note="assembly\_fragment"  
10097. 12552  
/note="assembly\_fragment"  
12653. 20026  
/note="assembly\_fragment"  
20127. 27970  
/note="assembly\_fragment"  
28071. 36727  
/note="assembly\_fragment"  
36828. 50972  
/note="assembly\_fragment"  
51073. 68989  
/note="assembly\_fragment"  
69090. 91875  
/note="assembly\_fragment"  
91976. 129030  
/note="assembly\_fragment"  
129131. 162409  
/note="assembly\_fragment"  
clone\_end:TY  
vector\_side:right  
BASE COUNT 47426 a 31849 c 32034 g 49990 t 1110 others  
ORIGIN

Query Match 24.0%; Score 80.6; DB 67; Length 162409;  
Best Local Similarity 77.7%; Pred. No. 1.6e-15;  
Matches 146; Conservative 0; Mismatches 37; Indels 5; Gaps 4;

QY 2 GCTGATCTTTTCTATGCTCTCCGCGCGCTGATGAGATACAGACAGAAACG 61  
DB 106888 GCTGATCTTTTCTATGCTCTCCGCGCGATATGAGATACAGACAGTAAACG 106947  
QY 62 GACAGCTGCTTTCATGATGATGATGATGATGATGATGATGATGATGATGATG 121  
DB 106948 GACAGCTCAT-TCATATATGATGATGATGATGATGATGATGATGATGATG 107004  
QY 122 NATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 181  
DB 107005 GATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 107062  
QY 182 TCTGTGAG 189  
DB 107063 GAGGAGAG 107070

RESULT 3  
HSNDPG 1846 bp mRNA PRI 03-DEC-1993  
LOCUS H.sapiens mRNA NDP.  
DEFINITION X65882.1 S61557  
ACCESSION X65882.1 GI:35016  
VERSION NDP gene.  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1846)  
AUTHORS Chen, Z.  
TITLE Direct Submission  
JOURNAL Submitted (28-APR-1992) Z. Chen, University of Oxford, Genetics  
Lab/Dept of Biochemistry, South Parks Road, Oxford OX1 3QU, UK  
REFERENCE 2 (bases 1 to 1846)  
AUTHORS Chen, Z.Y., Hendriks, R.W., Jobling, M.A., Powell, J.F.,  
Breakfield, X.O., Sims, K.B. and Craig, I.W.



```

source
1. .1769
/organism="Mus musculus"
/strain="1129"
/db_xref="taxon:10090"
/chromosome="X"
/dev_stage="embryonic stem cells"
/cell_type="ES cells"
/clone_lib="FIXII phage library"
429. .824
/gene="NDP"
429. .824
/gene="NDP"
429. .824
/gene="NDP"
/feature="Norrie disease locus"
/codon_start=1
/protein_id="CAA58725.1"
/db_xref="GI:854319"
/db_xref="MGI:102570"
/db_xref="SWISS-PROT:P48744"
/translation="MRNHVLAASISMLSLAIMGDTDSKTSSEFLMDSORCMRHHYVD
SISHPLYKSSKMYLLARCEGHCQSQRSEPLVSRSTVLKQPFRRSCHCRPQTSKTK
ALRLRCSGMRLATRYIILSCHECCSS"
1744. .1749
polya_signal
BASE COUNT 444 a 431 c 385 g 509 t
ORIGIN

```

```

Query Match
Best Local Similarity 17.6%; Score 59.2; DB 94; Length 1774;
Matches 64; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```

```

QY 2 GCTGCATCTTTTCATGCTCTCCCTGCTGCGCTGATGGAGATACAGACAGCAAAACG 61
DB 447 GCTGCATCTTTTCATGCTCTCCCTGCTGCGCTGATGGAGATACAGACAGCAAAACG 61
QY 62 GACAGCTCTGTC 74
DB 507 GACAGTTCATTTC 519

```

```

RESULT 6
LOCUS MNORRIE 1774 bp mRNA ROD 09-FEB-1996
DEFINITION M.musculus mRNA for Norrie disease gene product.
ACCESSION X92397
VERSION X92397.1 GI:1177544
KEYWORDS ND gene; Norrie disease locus.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 1774)
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE Hamelster, H., Wieringa, B., Hendriks, W., and Ropers, H.H.
An animal model for Norrie disease (ND): gene targeting of the
mouse ND gene
Hum. Mol. Genet. 5 (1), 51-59 (1996)
JOURNAL 2 (bases 1 to 1774)
REFERENCE 1 (bases 1 to 1774)
AUTHORS Berger, W.
TITLE Direct Submission
JOURNAL Submitted (17-Oct-1995) W. Berger, MPI fuer Molekulare Genetik,
Imnestrasse 73, D-14195 Berlin, FRG
FEATURES
source
1. .1774
/organism="Mus musculus"
/strain="C57/B16"
/db_xref="taxon:10090"
/chromosome="X"
/dev_stage="19 weeks"
/tissue_type="brain"
1. .225
/number=1
226. .599
exon
exon

```

```

gene
/number=2
join(432. .599, 600. .827)
/gene="ND"
join(432. .599, 600. .827)
CDS
/gene="ND"
/codon_start=1
/product="Norrie disease gene product"
/protein_id="CAA63134.1"
/db_xref="GI:1177545"
/db_xref="MGI:102570"
/db_xref="SWISS-PROT:P48744"
/translation="MRNHVLAASISMLSLAIMGDTDSKTSSEFLMDSORCMRHHYVD
SISHPLYKSSKMYLLARCEGHCQSQRSEPLVSRSTVLKQPFRRSCHCRPQTSKTK
ALRLRCSGMRLATRYIILSCHECCSS"
600. .1774
exon
/number=3
polya_signal
BASE COUNT 440 a 437 c 387 g 510 t
ORIGIN

```

```

Query Match
Best Local Similarity 17.6%; Score 59.2; DB 94; Length 1774;
Matches 64; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```

```

QY 2 GCTGCATCTTTTCATGCTCTCCCTGCTGCGCTGATGGAGATACAGACAGCAAAACG 61
DB 450 GCTGCATCTTTTCATGCTCTCCCTGCTGCGCTGATGGAGATACAGACAGCAAAACG 61
QY 62 GACAGCTCTGTC 74
DB 510 GACAGTTCATTTC 522

```

```

RESULT 7
LOCUS AC087621 63739 bp DNA HTG 15-JAN-2001
DEFINITION Homo sapiens chromosome 8 clone RP11-300E4 map 8, LOW-PASS SEQUENCE
ACCESSION AC087621
VERSION AC087621.1 GI:12229396
KEYWORDS HTG; HTGS_PHASED.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 63739)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE Birren, B., Linton, L., Nusbaum, C., and Lander, E.
Homo sapiens chromosome 8, clone RP11-300E4
JOURNAL 2 (bases 1 to 63739)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
Barn, N., Bastien, V., Boguslavsky, L., Boukhvalter, B., Brown, A.,
Cammarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,
Gardyna, S., Glade, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagob, B., Heath, A., Horton, L., Huine, W., Iliev, T., Johnson, R.,
Jones, C., Karatas, A., Labrecque, K., Lamazares, R., Landers, T.,
Lehoczky, J., Levine, R., Liu, G., Maclean, C., Macdonald, P.,
Margulis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
McPheters, R., Meldrum, J., Menus, L., Milova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Polara, V., Raymond, C., Retta, R.,
Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Roselli, M.,
Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P.,
Sougné, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Travers, M., Travis, N., Triglio, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zainoun, J.,
Zemek, L., Zimmer, A. and Zody, M.
Direct Submission
TITLE

```

JOURNAL  
Submitted (15-JAN-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MIBR

Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)

Project Information  
Center project name: L11683  
Center clone name: 300\_E\_4

NOTE: This record contains 79 individual  
sequencing reads that have not been assembled into  
contigs. Runs of N are used to separate the reads  
and the order in which they appear is completely  
arbitrary. Low-pass sequence sampling is useful for  
identifying clones that may be gene-rich and allows  
overlap relationships among clones to be deduced.  
However, it should not be assumed that this clone  
will be sequenced to completion. In the event that  
the record is updated, the accession number will  
be preserved.

1 700: contig of 700 bp in length  
\* 701 800: gap of 100 bp  
\* 801 1491: contig of 691 bp in length  
\* 1492 1591: gap of 100 bp  
\* 1592 2304: contig of 713 bp in length  
\* 2305 2404: gap of 100 bp  
\* 2405 3109: contig of 705 bp in length  
\* 3110 3209: gap of 100 bp  
\* 3210 3926: contig of 717 bp in length  
\* 3927 4026: gap of 100 bp  
\* 4027 4745: contig of 719 bp in length  
\* 4746 4845: gap of 100 bp  
\* 4846 5573: contig of 728 bp in length  
\* 5574 5673: gap of 100 bp  
\* 5674 6385: contig of 722 bp in length  
\* 6386 6495: gap of 100 bp  
\* 6496 7204: contig of 709 bp in length  
\* 7205 7304: gap of 100 bp  
\* 7305 8014: contig of 710 bp in length  
\* 8015 8114: gap of 100 bp  
\* 8115 8828: contig of 714 bp in length  
\* 8829 8928: gap of 100 bp  
\* 8929 9622: contig of 694 bp in length  
\* 9623 9722: gap of 100 bp  
\* 9723 10431: contig of 709 bp in length  
\* 10432 10531: gap of 100 bp  
\* 10532 11230: contig of 699 bp in length  
\* 11231 11330: gap of 100 bp  
\* 11331 11970: contig of 640 bp in length  
\* 11971 12070: gap of 100 bp  
\* 12071 12823: contig of 753 bp in length  
\* 12824 12923: gap of 100 bp  
\* 12924 13630: contig of 707 bp in length  
\* 13631 13730: gap of 100 bp  
\* 13731 14448: contig of 718 bp in length  
\* 14449 14548: gap of 100 bp  
\* 14549 15262: contig of 714 bp in length  
\* 15263 15362: gap of 100 bp  
\* 15363 16076: contig of 714 bp in length  
\* 16077 16176: gap of 100 bp  
\* 16177 16880: contig of 704 bp in length  
\* 16881 16980: gap of 100 bp  
\* 16981 17717: contig of 737 bp in length  
\* 17718 17817: gap of 100 bp  
\* 17818 18522: contig of 705 bp in length  
\* 18523 18622: gap of 100 bp  
\* 18623 19343: contig of 721 bp in length  
\* 19344 19443: gap of 100 bp

19444 20161: contig of 718 bp in length  
\* 20162 20261: gap of 100 bp  
\* 20262 20983: contig of 722 bp in length  
\* 20984 21083: gap of 100 bp  
\* 21084 21810: contig of 727 bp in length  
\* 21811 21910: gap of 100 bp  
\* 21911 22637: contig of 727 bp in length  
\* 22638 22737: gap of 100 bp  
\* 22738 23439: contig of 702 bp in length  
\* 23440 23539: gap of 100 bp  
\* 23540 24247: contig of 708 bp in length  
\* 24248 24347: gap of 100 bp  
\* 24348 25051: contig of 704 bp in length  
\* 25052 25151: gap of 100 bp  
\* 25152 25851: contig of 700 bp in length  
\* 25852 25951: gap of 100 bp  
\* 25952 26675: contig of 724 bp in length  
\* 26676 26775: gap of 100 bp  
\* 26776 27470: contig of 695 bp in length  
\* 27471 27570: gap of 100 bp  
\* 27571 28264: contig of 694 bp in length  
\* 28265 28364: gap of 100 bp  
\* 28365 29085: contig of 721 bp in length  
\* 29086 29185: gap of 100 bp  
\* 29186 29892: contig of 707 bp in length  
\* 29893 29992: gap of 100 bp  
\* 29993 30703: contig of 711 bp in length  
\* 30704 30803: gap of 100 bp  
\* 30804 31509: contig of 706 bp in length  
\* 31510 31609: gap of 100 bp  
\* 31610 32297: contig of 688 bp in length  
\* 32298 32397: gap of 100 bp  
\* 32398 33104: contig of 707 bp in length  
\* 33105 33204: gap of 100 bp  
\* 33205 33905: contig of 701 bp in length  
\* 33906 34005: gap of 100 bp  
\* 34006 34708: contig of 703 bp in length  
\* 34709 34808: gap of 100 bp  
\* 34809 35513: contig of 705 bp in length  
\* 35514 35613: gap of 100 bp  
\* 35614 36345: contig of 732 bp in length  
\* 36346 36445: gap of 100 bp  
\* 36446 37174: contig of 729 bp in length  
\* 37175 37274: gap of 100 bp  
\* 37275 37987: contig of 713 bp in length  
\* 37988 38087: gap of 100 bp  
\* 38088 38800: contig of 713 bp in length  
\* 38801 38900: gap of 100 bp  
\* 38901 39598: contig of 698 bp in length  
\* 39599 39698: gap of 100 bp  
\* 39699 40403: contig of 705 bp in length  
\* 40404 40503: gap of 100 bp  
\* 40504 41221: contig of 718 bp in length  
\* 41222 41321: gap of 100 bp  
\* 41322 42007: contig of 686 bp in length  
\* 42008 42107: gap of 100 bp  
\* 42108 42794: contig of 687 bp in length  
\* 42795 42894: gap of 100 bp  
\* 42895 43612: contig of 718 bp in length  
\* 43613 43712: gap of 100 bp  
\* 43713 44429: contig of 717 bp in length  
\* 44430 44529: gap of 100 bp  
\* 44530 45239: contig of 710 bp in length  
\* 45240 45339: gap of 100 bp  
\* 45340 46049: contig of 710 bp in length  
\* 46050 46149: gap of 100 bp  
\* 46150 46864: contig of 715 bp in length  
\* 46865 46964: gap of 100 bp  
\* 46965 47674: contig of 710 bp in length  
\* 47675 47774: gap of 100 bp  
\* 47775 48465: contig of 691 bp in length  
\* 48466 48565: gap of 100 bp  
\* 48566 49271: contig of 706 bp in length

```

* 49272 49371: gap of 100 bp
* 49372 50018: contig of 647 bp in length
* 50018 50118: gap of 100 bp
* 50118 50839: contig of 721 bp in length
* 50839 50939: gap of 100 bp
* 50939 51664: contig of 725 bp in length
* 51664 51764: gap of 100 bp
* 51764 52477: contig of 713 bp in length
* 52477 52577: gap of 100 bp
* 52577 53281: contig of 704 bp in length
* 53281 53381: gap of 100 bp
* 53381 54081: contig of 700 bp in length
* 54081 54181: gap of 100 bp
* 54181 54877: contig of 696 bp in length
* 54877 54977: gap of 100 bp
* 54977 55682: contig of 705 bp in length
* 55682 55782: gap of 100 bp

Query Match
  12.2% Score 41; DB 77; Length 63739;
  52.5% Pred. No. 0.019;
Matches 74; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 173 TGGGAGTCTTCGTGAGACCTATGATGCCCGGCGCCCTGAGNTCCAGAGGAGNC 232
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 51985 TGGGCCCTTCTCCCTGCTCTTCTCCATCCCGAGGTGTGGCAGGTATMAAGGAGAGA 52044
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 233 TGTGCTCCCATCGCTTACAGAGGTTCAAGNAAGGGCCCGCCATGCCAGTCCCTTGN 292
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 52045 TGTATGGAACGTCCCTCCAGGAGAGGAGGAGGAGGCCCGCTGACATCAGCGCTTGT 52104
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 293 CAGNAGNANGANTGTGNC 313
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 52105 TGGGAGCCACAGACAGGACC 52125
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
AC004736/c
LOCUS
DEFINITION
AC004736 115958 bp DNA PRI 03-JUL-1998
Human chromosome 11p14.3 PAC clone PDJ1082L12 containing KNCN1 and
AC004736 complete sequence.
AC004736
VERSION
AC004736.1 GI:3282183
KEYWORDS
HTG.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
1 (bases 1 to 115958)
Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basil,M.,
Buettner,J., Bumeister,R., Card,P., desaillobat,F., Dunn,J.,
English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G.,
Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N.,
Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S.,
Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.
HTGS Submission
Unpublished
2 (bases 1 to 115958)
Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basil,M.,
Buettner,J., Bumeister,R., Card,P., desaillobat,F., Dunn,J.,
English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G.,
Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N.,
McFarland,J., Miller,R., Newton,J., Osborne-Lawrence,S.,
Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.
Direct Submission
Submitted (23-MAY-1998) Genome Science & Technology Center,
University of Texas Southwestern Medical Center, 5323 Harry Hines
Blvd, Dallas, TX 75235-8591, USA
3 (bases 1 to 115958)
Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basil,M.,
Buettner,J., Bumeister,R., Card,P., desaillobat,F., Dunn,J.,
English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G.,
Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N.,
Major,T., McFarland,J., Harris,J., Lewis,E., Loo,H., Loo,K.N.,
Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S.,

```

```

TITLE
Direct Submission
JOURNAL
Submitted (01-JUL-1998) Genome Science & Technology Center,
University of Texas Southwestern Medical Center, 5323 Harry Hines
Blvd, Dallas, TX 75235-8591, USA
4 (bases 1 to 115958)
Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basil,M.,
Buettner,J., Bumeister,R., Card,P., desaillobat,F., Dunn,J.,
English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G.,
Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N.,
Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S.,
Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.
Direct Submission
Submitted (03-JUL-1998) Genome Science & Technology Center,
University of Texas Southwestern Medical Center, 5323 Harry Hines
Blvd, Dallas, TX 75235-8591, USA
On Jul 1, 1998 this sequence version replaced gi:3152625.
NOTE: Extent of PDJ1082L12 overlap with mapped clones: SP6 PAC
cloning end: 1..8775, PDJ59M18; T7 PAC cloning end: 45252..115958,
6-109h6. Further information regarding the map of this region or
annotation of PDJ1082L12 can be found at
http://gestec.smed.edu/chromosome2.htm.
CHROMOSOMAL LOCUS: This PAC clone comes from the Usher syndrome
MARKER CONFIRMATION: STS/STS sequence confirmed; D11S4130, KNCN1,
and MYO1.
MAILED CLONE OVERLAP: PDJ59M18 and 6-109h6.
IMPORTANT: This submission contains the entire insert of clone
PDJ1082L12. PDJ1082L12 comes from a PAC library constructed at the
Roswell Park Cancer Institute by the Pletier de Jong group. This
clone has been finished according to strict quality criteria and
attempts have been made to resolve all base calling problems such
as compressions and repetitive elements. The expected Phred/Phrap
calculated errors/10kb is 0.39. In addition, this sequence has
been finished such that 99.9% of consensus base calls consist of
either double-stranded coverage or 2 types of labeling chemistry on
one strand.
FEATURES
location/qualifiers
1..115958
/organism="Homo sapiens"
/db_xref="taxon:9606"
4409..4517
/rpl_family="MIR"
complement(19171..19291)
/rpl_family="Alu"
complement(21561..21985)
/rpl_family="Alu"
23678..23767
/rpl_family="MIR"
28877..31250
/rpl_family="Tiggr1"
complement(31777..31876)
/rpl_family="MIR"
complement(36023..36231)
/rpl_family="MIR"
complement(39712..39766)
/rpl_family="MIR"
complement(41971..42073)
/rpl_family="MIR"
45642..45752
/rpl_family="MIR"
57204..57337
/rpl_family="MIR"
complement(58027..58249)
/rpl_family="MIR20"
59316..59729
/rpl_family="L1"
complement(60013..60386)
/rpl_family="RHE1"
60398..60596
/rpl_family="L1"
60940..61228
/rpl_family="Alu"

```



```
repeat_region complement(63635, .63900)
                /rpt_family="Alu"
repeat_region complement(75510, .75774)
                /rpt_family="Alu"
repeat_region complement(82106, .82378)
                /rpt_family="Alu"
repeat_region complement(84150, .84211)
                /rpt_family="MLT1"
repeat_region 84727, .85172
                /rpt_family="MLT1"
repeat_region 85852, .86134
                /rpt_family="Alu"
repeat_region 86387, .86565
                /rpt_family="Alu"
repeat_region complement(88437, .88514)
                /rpt_family="MIR"
repeat_region complement(90686, .90752)
                /rpt_family="MIR"
repeat_region 91042, .91163
                /rpt_family="MIR"
repeat_region 94274, .94419
                /rpt_family="MIR"
repeat_region complement(95470, .95617)
                /rpt_family="MLT1"
repeat_region complement(97484, .97917)
                /rpt_family="L1"
repeat_region 105586, .105657
                /rpt_family="L1"
repeat_region 105673, .105884
                /rpt_family="MIR"
repeat_region /rpt_family="MIR"
                /rpt_family="MIR"
repeat_region complement(106369, .106641)
                /rpt_family="Alu"
repeat_region complement(107302, .107596)
                /rpt_family="Alu"
repeat_region 109188, .109458
                /rpt_family="Alu"
repeat_region /rpt_family="Alu"
                complement(110473, .110773)
repeat_region /rpt_family="Alu"
                complement(111227, .111337)
repeat_region /rpt_family="L1"
                111412, .111792
repeat_region /rpt_family="THE1"
                complement(112629, .112992)
repeat_region /rpt_family="L1"

BASE COUNT 28386 a 29554 c 29450 g 28568 t
ORIGIN
```

Query Match 12.2% Score 41; DB 85; Length 115958;  
Best Local Similarity 52.5% Pred. NO. 0.019;  
Matches 74; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

```
Oy 173 TGGGAGTTTCTGTGAGACATCTGATCCCGGNNTGCCCTGAGNTCCAGAGGNC 232
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41574 TGGGCGCTTGTCTCCCTGCTTCCTCCATCCCGAGGTGCGGTAAAGGGAGGA 41515

Oy 233 TGTCTGCTCCATCGCTCGCTGAGGCGCCGCMATGCGAGTCCCTTGGN 292
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41514 TGATGAGACGTCGCTCCAGAGGAGAGGAGGAGGAGGCGCTGACATCAGCCTTGGT 41455

Oy 293 CAGNAGNANGANTTGNC 313
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41454 TGGGAGCCAGCACAGGAGCC 41434
```

RESULT 9  
AC010136/c 171453 bp DNA HTG 22-MAR-2001  
LOCUS Homo sapiens chromosome UNK clone RP11-129D2, WORKING DRAFT  
DEFINITION  
SEQUENCE 2 unordered pieces.  
AC010136  
AC010136.7 GI:13431097  
VERSION  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEPIN.

SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 171453)  
AUTHORS Waterston, R.H.  
TITLE The sequence of Homo sapiens clone  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 171453)  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (13-SEP-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Mar 22, 2001 this sequence version replaced gi:11990739.

## COMMENT

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
----- Project Information -----  
Center project name: H.NH0129D02  
----- Summary Statistics -----  
Sequencing vector: p13, 588  
Sequencing vector: plasmid, 428  
Chemistry: Dye-Primer ET; 52% of reads  
Chemistry: Dye-terminator Big Dye; 48% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 169816 bases at least Q40  
Consensus quality: 170375 bases at least Q20  
Consensus quality: 170797 bases at least Q20  
Insert size: 171353; sum-of-contigs  
Quality coverage: 7.08 in Q20 bases; agarose-fp  
Quality coverage: 7.13 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 33328: contig of 33328 bp in length  
\* 33329 32428: gap of unknown length  
\* 32429 171453: contig of 139025 bp in length.

## FEATURES

## source

1. 171453  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="UNK"  
/clone="RP11-129D2"

## misc\_feature

1. 33328  
/note="assembly\_name:Contig7  
clone\_end:T7  
vector\_side:left"  
32429, .171453  
/note="assembly\_name:Contig8  
clone\_end:SP6  
vector\_side:right"

## misc\_feature

BASE COUNT 38981 a 41426 c 46891 g 44055 t 100 others  
ORIGIN

Query Match 10.4% Score 34.8; DB 61; Length 171453;  
Best Local Similarity 52.6% Pred. NO. 2.2;  
Matches 60; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

```
Oy 202 CCCCNNNTGGCCCTGAGTTCAGAGGNCCTGCGTCCCTGCGCTGAGGAGGTTCAA 261
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1082 CTTGATTGGCCATTTCCTCCAGAGGGGCGAGAGGAGGAGGAGGCTTGTGCTGGGAA 1023
```





```

* 2372 3377: contig of 1006 bp in length
* 3378 3477: gap of 100 bp
* 3478 4831: contig of 1354 bp in length
* 4832 4931: gap of 100 bp
* 4932 6003: contig of 1072 bp in length
* 6004 6103: gap of 100 bp
* 6104 7238: contig of 1135 bp in length
* 7239 7338: gap of 100 bp
* 7339 8789: contig of 1451 bp in length
* 8790 8889: gap of 100 bp
* 8890 10365: contig of 1476 bp in length
* 10366 10465: gap of 100 bp
* 10466 11651: contig of 1186 bp in length
* 11652 11751: gap of 100 bp
* 11752 13621: contig of 1870 bp in length
* 13622 13721: gap of 100 bp
* 13722 15643: contig of 1922 bp in length
* 15644 15743: gap of 100 bp
* 15744 17047: contig of 1304 bp in length
* 17048 17147: gap of 100 bp
* 17148 18830: contig of 1683 bp in length
* 18831 18930: gap of 100 bp
* 18931 20799: contig of 1869 bp in length
* 20800 20899: gap of 100 bp
* 20900 25005: contig of 4106 bp in length
* 25006 25105: gap of 100 bp
* 25106 31884: contig of 6779 bp in length
* 31885 31984: gap of 100 bp
* 31985 44283: contig of 12289 bp in length
* 44284 44383: gap of 100 bp
* 44384 55405: contig of 11022 bp in length
* 55406 55505: gap of 100 bp
* 55506 71682: contig of 16177 bp in length
* 71683 71782: gap of 100 bp
* 71783 97873: contig of 26091 bp in length
* 97874 97973: gap of 100 bp
* 97974 127625: contig of 29652 bp in length
* 127626 127725: gap of 100 bp
* 127726 180359: contig of 52634 bp in length.

```

## FEATURES

source

```

1.180359
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-296G21"
/clone.lib="RPC1-11 Human Male BAC"
1.1137
/note="assembly_fragment"
misc_feature
1238..2271
/note="assembly_fragment"
misc_feature
2372..3377
/note="assembly_fragment"
misc_feature
3478..4831
/note="assembly_fragment"
misc_feature
4932..6003
/note="assembly_fragment"
misc_feature
6104..7238
/note="assembly_fragment"
misc_feature
7339..8789
/note="assembly_fragment"
misc_feature
8890..10365
/note="assembly_fragment"
misc_feature
10466..11651
/note="assembly_fragment"
misc_feature
11752..13621
/note="assembly_fragment"
misc_feature
13722..15643
/note="assembly_fragment"
misc_feature
15744..17047
/note="assembly_fragment"
misc_feature
17148..18830
/note="assembly_fragment"

```

```

misc_feature 18931..20799
/note="assembly_fragment"
misc_feature 20900..25005
/note="assembly_fragment"
misc_feature 25106..31884
/note="assembly_fragment"
misc_feature 31985..44283
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
4384..55405
/note="assembly_fragment"
55506..71682
/note="assembly_fragment"
71783..97873
/note="assembly_fragment"
clone_end:T7
vector_side:right"
97974..127625
/note="assembly_fragment"
127726..180359
/note="assembly_fragment"
BASE COUNT 48627 a 42242 c 41192 g 46193 t 2105 others
ORIGIN

```

```

Query Match 9.8%; Score 33; DB 67; Length 180359;
Best Local Similarity 60.0%; Pred. No. 8.5;
Matches 54; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

```

```

QY 2 GCTGCACTTTTTCATGCTCTCCCTGCGCCTATGAGACATACAGACAAACG 61
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 161279 GTTTCATGTTCTTGACACTTCCCTCTGGAGACAGTGAAGGGTCCACAGACTGACCA 161220
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 62 GACAGCTGCTGTCATGATGATCGATCGAGCC 91
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 161219 GACAGCTGACAGCTGTCACAGACGCTCAC 161190
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 13
AC006059/C AC006059 185161 bp DNA PRI 01-APR-1999
LOCUS Homo sapiens PAC RPC14-613B23 (Roswell Park Cancer Institute Human
DEFINITION PAC Library) complete sequence.
ACCESSION AC006059
VERSION AC006059.3 GI:4544348
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 185161)
Muzny,D., Areson,A.D., Bouck,J., Bunac,C., Chen,J., Chen,Z.,
Culpeper,P., Ding,Y., Dugan,S.P., Durbin,K.J., Forcum,J.,
Ganesh,R.P., Garcia,C., Garcia,D.K., Gorrell,H., Gorrell,L.L.,
He,X., Hernandez,J., Jackson,L.E., Kondrjewski,N., Leal,B.,
Lichtarge,O., Liu,W., Logan,O., Lu,J., Martinez,C., Moore,S.,
Moorish,T., Nguyen,N., Oswal,G., Pampell,L.R., Parish,B.J.,
Perez,L.M., Rashid,N.D., Rives,C.M., Scherer,S.E., Shen,H.,
Simion,M.L., Vo,O.K., Wei,Y., Williamson,A.L., Worley,K., Zhou,X.,
Naylor,S.L. and Gibbs,R.A.
Direct Submission
2 (bases 1 to 185161)
Unpublished
Worley,K.C.
Direct Submission
Submitted (26-NOV-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 185161)
Worley,K.C.
Direct Submission
Submitted (30-MAR-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

```

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

## COMMENT

Baylor Plaza, Houston, TX 77030, USA  
4 (bases 1 to 185161)  
Morley, K.C.  
Direct Submission  
Submitted (01-APR-1999) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Mar 30, 1999 this sequence version replaced gl:4263627.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
[gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the  
entire insert of this clone. Overlapping regions of clones are only  
sequenced and submitted once, so the sequence for the remainder of  
the insert may be found in the record for the adjacent clones.  
Overlapping clones are noted at the beginning and end of the  
Features listing.

## ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches  
of a local database that includes entries from dbSTS, GDB, and  
local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,  
unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST  
(Nuc. Acids Res. 25:3389-3402) similarly (expect < 1e-34) to the  
EST and cDNA sequences. Genes demonstrate at least two exons  
flanked by consensus splice sites that maintained sequence  
continuity across the splice junctions. Sequences that are not  
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
standard of double strand coverage with a minimum of 2 clones and 2  
reads with no ambiguities or 2 chemistries with a minimum of 2  
clones and 3 reads with no ambiguities. If the sequence quality for  
a region does not meet this standard, it will be indicated in the  
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
standards - estimated error rate less than 1 per 10,000 bases.  
Reports of lowest quality individual bases and measures of base  
quality are listed below. Description of the metrics can be found  
at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

## QUALSTAT-REPORT-----

----- Summary Statistics -----  
Contig Length: 185161  
Phrap values in estimate: 183812  
Average error rate (BCM-Phrap estimate): 0.000104921  
Fraction of Phrap values less than 40 : 0.0374023  
Number of consensus changing edits: 20  
Number of N's in consensus : 0

----- Consensus changing edits -----  
Position Original+Context Edited+Context  
28864 attcaggcag(l)tcgagagcca attcaggcag(g)gcgagagcca  
28865 attcaggcag(l)tcgagagcag attcaggcag(g)gcgagagcag  
32674 ctcaagcgag(n)aacagcctag ctcaagcgag(c)aacagcctag  
53006 caccgtgcca(a)gctattttt caccgtgcca(g)gctattttt  
84131 ctcccaaac(n)ctcccaacc ctcccaaac(c)ctcccaacc  
100219 ctaggaaact(n)cactcccttc ctaggaaact(g)cactcccttc  
109068 tccctcctcct(n)ncctccttc tccctcctcct(c)ccctccttc  
109069 cccctcctc(n)ccctccttc cccctcctc(c)ccctccttc  
149456 ctctctctcc(n)ggaaaaaaa ctctctctcc(l)ggaaaaaaa  
149636 aataaataat(n)caagagcaga aataaataat(c)caagagcaga  
158518 ttacaagta(n)aaaagagcat ttacaagta(a)aaaagagcat  
153739 tgacatccag(n)atccagcct tgacatccag(c)atccagcct  
167565 gagggcgagg(n)tgagacgaag gagggcgagg(g)tgagacgaag  
169144 ttgtgtcca(n)taactgtgt ttgtgtcca(a)tgatctgtgt

169146 ttgttccat(n)atctgtctc ttgttccat(g)atctgtctc  
169213 aacattacatg(n)ncgggcgttg aacattacatg(g)ncgggcgttg  
169214 cattacatg(n)cgggcgttgt cattacatg(g)cgggcgttgt  
176105 ctacaataat(n)aaaattagca ctacaataat(a)aaaattagca  
178686 gaattgttt(c)gtggaagaca gaattgttt(c)gtggaagaca  
183324 gttgttgtt(n)ttgttttgt gttgttgtt(t)ttgttttgt

## ----- Distribution of Quality &lt; 40 Bases -----

# bases	5	10	15	20	25	30	35	40
10001							*	*
9001							*	*
8001							*	*
7001							*	*
6001							*	*
5001					*	*	*	*
4001				*	*	*	*	*
3001			*	*	*	*	*	*
2001		*	*	*	*	*	*	*
1001		*	*	*	*	*	*	*
01	*	*	*	*	*	*	*	*

## Version: 1.01 gxf0.

FEATURES  
Source Location/Qualifiers

1..185161  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RPC14-613B23"  
complement(1..79)  
/rpt\_family="AluS"  
complement(80..484)  
/rpt\_family="L1M4"  
complement(485..785)  
/rpt\_family="AluSg"  
complement(786..1830)  
/rpt\_family="L1M4"  
complement(1851..1932)  
/rpt\_family="AluJ/FRAM"  
2846..2885.  
/rpt\_family="MER5B"  
2886..3070  
/rpt\_family="AluSp"  
complement(3263..3380)  
/rpt\_family="MIR"  
4087..4629  
/rpt\_family="MER4B"  
4659..4970  
/rpt\_family="AluSx"  
5120..5508  
/rpt\_family="MER57B"  
complement(5509..5687)  
/rpt\_family="AluJb"  
5712..6160  
/rpt\_family="MER77"  
6295..6461  
/rpt\_family="MER5B"  
complement(6858..7224)  
/rpt\_family="L2"  
7524..7663  
/rpt\_family="L1M1"  
complement(7666..7699)  
/rpt\_family="AT-rich"  
complement(7700..7790)  
/rpt\_family="AluSg/x"  
complement(7791..7832)  
/rpt\_family="AluSc"  
complement(7831..7876)  
/rpt\_family="AluSg/x"

```

repeat_region complement(7877..8139)
repeat_region /rpt_family="AluSx"
repeat_region complement(8140..8161)
repeat_region /rpt_family="(TAA)n"
repeat_region complement(8162..8463)
repeat_region /rpt_family="AluSq"
repeat_region 8464..9146
repeat_region /rpt_family="L1ME1"
repeat_region complement(9148..9436)
repeat_region /rpt_family="AluDb"
repeat_region 9487..9712
repeat_region /rpt_family="L1ME3"
repeat_region complement(9715..9841)
repeat_region /rpt_family="MER5B"
repeat_region complement(10378..10411)
repeat_region /rpt_family="(Ca)n"
repeat_region 10424..10497
repeat_region /rpt_family="L1M4"
repeat_region complement(10482..10751)
repeat_region /rpt_family="L1M4"
repeat_region complement(10753..11023)
repeat_region /rpt_family="AluDo"

Query Match 9.8%; Score 33; DB 86; Length 185161;
Best Local Similarity 60.0%; Pred. No. 8.5;
Matches 54; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 2 GCAGCATCTTTTCTATGCTCTCCCTGCGCTGATGGAGATACAGACGAAACG 61
Db 101696 GTTTCATGTTCTTGTACACTCCCTCTGGAGTGAAGTGGGCTCCAGACGCTGACCA 101637
QY 62 GACAGCTGCTGTCATGTCATGTCGACCTGACCC 91
Db 101636 GACAGCTGTCGACCTGTCATGACAGCGTCAC 101607

RESULT 14
LOCUS AC073530/c
DEFINITION Homo sapiens chromosome 12 clone RP11-123010. WORKING DRAFT
ACCESSION AC073530
VERSION AC073530.17 GI:13162369
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 161547)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alstrooms,S.L., Amaralungu,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Blumh,K., Blankenburg,K., Bonini,D., Bouck,J.,
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C.,
Burich,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Cartier,M., Chavez,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,X., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Barnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,S., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hoggins,M., Hollway,C.,
Hollins,B., Homsi,F., Howard,S., Huber,J., Huylk,S., Hume,J.,
Joudan,S., Karlsson,E., Kelly,S., Johnson,R., Jolivet,S.,
Kovach,S., Kratovic,J., Kuresh,A., Landy,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseged,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,J., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,

```

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

```

Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: HBOR
Center clone name: RP11-123010
Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-Primer Bodypy: 15% of reads
Chemistry: Dye-terminator Big Dye: 85% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 160458 bases at least Q40
Consensus quality: 161972 bases at least Q20
Consensus quality: 163257 bases at least Q20
Estimated insert size: 157486; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 8.9x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 156390: contig of 156390 bp in length
* 156391 156490: gap of unknown length
* 156491 157492: contig of 1002 bp in length
* 157493 157592: gap of unknown length
* 157593 158938: contig of 1346 bp in length
* 158939 159038: gap of unknown length
* 159039 160260: contig of 1222 bp in length
* 160261 160360: gap of unknown length
* 160361 161547: contig of 1187 bp in length.
Location/Qualifiers
1. 161547
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-123010"

```

BASE COUNT  
ORIGIN

48691 a 31251 c 31335 g 49855 t 415 others

```

Query Match Similarity: 9.8%; Score 32.8; DB 74; Length 161547;
Best Local Alignment: 53.9%; Pred. No. 9.8;
Matches 55; Conservative 0; Mismatches 47; Indels 0; Gaps 0

Oy 224 AGAGAGNCCTGTCGTCGCCATCCGCTAGCAGGCTTCAGNAAGAGGCGCCGNCATGACA 283
      ||||| 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 79228 AGAGGTCACAGGCCCATTCGATATACCAAGCCTTGGTTTCACAGAAAGGTTCACTTAAGAT 79169
      ||||| 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

Oy 284 GTCTTGGNCAGNAGNANGANTTGGNCCCAACCCCTTGG 325
      ||||| 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 79168 AGACTTGGAGTGTAGACCAAGAGTACGAGAACAATACCTTGG 79127
      ||||| 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT 15
AC069409/c AC069409 177765 bp DNA HTG 06-SEP-2000.
DEFINITION Homo sapiens chromosome 3 clone RP11-308L22, WORKING DRAFT
SEQUENCE AC069409
AC069409 24 unordered pieces.
AC069409.7 GI:9966586
HTG: HTGS_PHA5E1; HTGS_DRAFT.
ORIGIN
ORIGIN human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 177765)
AUTHORS Muny,D.M., Adams,C., Bailly,M., Barbara,J., Blankenburg,K.,
Bodet,B., Bowck,J.J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
Forcum-Tansey,J.J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
Guaveira,W., Harris,K., Hernandez,J., Hodgson,A., Hogue,M.,
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
Kelly,S., Kondejowski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
Lichterage,O., Liu,J., Liu,W., Logan,O., Lozano,R.J., Lu,J.,
Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mel,G., Morgan,M.,
Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,C.,
Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sugand,R.,
Tabour,P., Taylor,T., Vasquez,L., Vinson,K., Vo,O., Wabhan,M.,
Watlington,S., Weinstein,G., Weinstein,I.R., Williamson,A.,
Worley,K., Wren,J., Wrenford,G., Yu,W., Zhou,X., Nelson,D. and
Gibbs,R.
Direct Submission
Unpublished 2 (bases 1 to 177765)
Worley,K.C.
Direct Submission
Submitted (30-MAY-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 5, 2000 this sequence version replaced gi:8844019.

COMMENT
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HAVK
Center clone name: RP11-308L22
----- Summary Statistics
Sequencing vector: M13; 108821
Chemistry: Dye-terminator Big Dye; 94% of reads
Assembly program: Phrap; versions 0.990329
Consensus quality: 156950 bases at least Q40
Consensus quality: 164968 bases at least Q30
Consensus quality: 170263 bases at least Q20
Estimated insert size: 169417; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-tp estimation
Quality coverage: 3.2x in Q20 bases; sum-of-contigs estimation

```

Db 37114 TCTGAACCTAGACGGTCCCATCTAGGGTGAAGAGACAGATCCTCAGGCATT 37055  
QY 68 TCGTTCATGATCGACTGGACCT 92  
Db 37054 AGATCTCATAGGAAGTCTGGACCT 37030

Search completed: August 24, 2001, 03:54:00  
Job time: 4216 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2001, 03:29:59 ; Search time 128.32 Seconds  
(without alignments)  
1644.131 Million cell updates/sec

Title: US-08-224-621-74

Perfect score: 336  
Sequence: 1 CGTCGACATCTTTCTATGTC.....CCCCNTGTGTTCCCAACCA 336

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 730101 seqs, 313950809 residues  
Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

N.Geneseq\_0601:\*  
1: /SIDSI/gcgdata/geneseq/geneseqn/NA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseqn/NA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseqn/NA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseqn/NA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseqn/NA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseqn/NA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseqn/NA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseqn/NA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseqn/NA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseqn/NA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseqn/NA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseqn/NA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseqn/NA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseqn/NA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseqn/NA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseqn/NA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseqn/NA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseqn/NA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseqn/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	9.8	4217	21	Human ORFX ORF2205
2	31.2	9.3	3067	15	Human EAA3c excita
3	30.6	9.1	1147	21	Arabidopsis thalia
4	30.6	9.1	1626	21	Arabidopsis thalia
5	30.4	9.0	5534	14	AAQ35988
6	30.2	9.0	965	20	AAQ20446
7	30.2	9.0	1275	21	AAA62025
8	30.2	9.0	1278	21	AAQ93406
9	30.2	9.0	1712	21	AAA62035
10	30.2	9.0	2171	21	AAQ09354
11	29.8	8.9	441	21	AAQ02302

C	12	29.8	8.9	15914	22	AAE57269
C	13	29.6	8.8	1160	18	AAE93294
C	14	29.6	8.8	1166	18	AAE93285
C	15	29.6	8.8	1169	18	AAE93282
C	16	29.6	8.8	1169	18	AAE93283
C	17	29.6	8.8	1169	18	AAE93284
C	18	29.6	8.8	1246	18	AAE93286
C	19	29.6	8.8	2602	18	AAE93286
C	20	29.4	8.8	14784	21	AAE64141
C	21	29.2	8.7	49999	20	AAE23901
C	22	28.6	8.5	2921	21	AAE30884
C	23	28.4	8.5	2773	20	AAE22704
C	24	28.4	8.5	6744	20	AAE38125
C	25	28.4	8.5	7286	17	AAE14547
C	26	28.4	8.5	7346	20	AAE77924
C	27	28.4	8.5	7560	22	AAE83437
C	28	28.2	8.4	849	21	AAE93855
C	29	28.2	8.4	3416	17	AAE30373
C	30	28.2	8.4	3416	19	AAE40859
C	31	28.2	8.4	3621	19	AAE17800
C	32	28.2	8.4	4089	21	AAE36693
C	33	28	8.3	1399	12	AAQ11843
C	34	28	8.3	3287	21	AAE98900
C	35	28	8.3	10732	21	AAE10594
C	36	28	8.3	34203	22	AAE74546
C	37	28	8.3	34263	22	AAE74547
C	38	28	8.3	534720	19	AAE30458
C	39	28	8.3	53615	19	AAE30459
C	40	27.8	8.3	49136	21	AAE27475
C	41	27.6	8.2	1332	21	AAE34344
C	42	27.4	8.2	2309	22	AAE67781
C	43	27.4	8.2	2834	20	AAE86274
C	44	27.4	8.2	2834	8	AAE70307
C	45	27.4	8.2	2834	10	AAE90394

ALIGNMENTS

RESULT	1	
AC	AACT6650	standard; cDNA; 4217 BP.
DT	08-FEB-2001	(first entry)
DE	Human ORFX ORF2205 polynucleotide sequence SEQ ID NO:4409.	
XX	Human: open reading frame; ORFX; detection; cytostatic; hepatotropic; vulnerable; antiproliferative; antiparkinsonian; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; candidant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antirheumatic; antihypertensive; antianemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive; ss.	
XX	Homo sapiens.	
OS	WO200058473-A2.	
PN	05-OCT-2000.	
PD	31-MAR-2000; 2000MO-US08621.	
PF	31-MAR-1999; 99US-0127607.	
XX		

Sequence 4217 BP; 888 A; 1243 C; 1230 G; 856 T; 0 other;

RESULT	2	
AAO81157/c		
ID	AAO81157	standard; cDNA; 3067 BP.
XX		
XX	AAO81157;	
AC		
XX		
DT	22-MAR-1995	(first entry)
XX		
DE	Human EAA3c excitatory amino acid receptor cDNA.	
XX		
KW	excitatory amino acid receptor; human EAA3 receptor; kainate-type	
KW	EAA3c receptor; ds.	
XX		
OS	Homo sapiens.	
XX		
XX		
key		Location/Qualifiers
FT		418..3015
CDS		/*tag= a
FT		

CA2110933-A.  
12-JUN-1994.  
08-DEC-1993; 93CA-2110933.  
11-DEC-1992; 92US-0989793.  
(KAMB/) KAMBOJ R.  
(ELLIOT) ELLIOT C E.  
(NUTT/) NUTT S L.  
ELLIOT CE, Kamboj R, Nutt SL,  
WPI; 1994-255829/32.  
P-PSDB; AAR63069.  
Polynucleotide encoding a human excitatory amino acid 3 receptor  
or fragment used to assay test ligands for their interaction  
with a human CBS receptor.

Claim 1, Fig 1 and Fig 4B, 35pp. English.

CC The human EAA3a receptor (AAR0112) and its variants EAA3b, EAA3c and  
CC EAA3d (AAR63068-R63070) occur naturally in human brain. They are all  
CC members of the EAA3 receptor family, having extracellular N- and C-  
CC termini and 4 internal hydrophobic domains which anchor the receptor  
CC within the cell surface membrane. The receptors bind glutamate and  
CC also exhibit binding properties characteristic of kainate-type EAA  
CC receptors.  
CC N.B. The cDNA and corresp. amino acid sequences for EAA3b, 3c and  
CC 3d do not appear in full in the specification; the sequences have  
CC been compiled by combining sequences in Figures 1 and 4 according  
CC to the description given in the disclosure.

Query Match	9.3%	Score 31.2	DB 15	Length 3067
Best Local Similarity	55.7%	Pred. No. 1.7		
Matches	54	Conservative	0	Mismatches 43
				Indels 0
				Gaps 0
QY	116	ATGTTGAAATTCATACGCTACCCCGTTGNTACAACTAGNCTCAAGAAGATAGTGG	175	
DB	3025	ATATGCAAACTTAACATACCACTCTCTAGAAAGATGTCATCCCTGTCAATATCAT	2966	
QY	176	GAGCTTCTGTGAGACCTATCTGAATCCCGGNNTGGC	212	
DB	2965	TATTCCTCGTATTTGATATGAATTTCCCAATAGC	2929	

RESULT	3
AAC36410/c	
ID	AAC36410 standard; DNA; 1147 bp.
XX	
XX	
AC	AAC36410;
XX	
DT	17-OCT-2000 (first entry)
XX	
De	Arabidopsis thaliana DNA fragment SEQ ID NO: 13699.
XX	
KW	Hybridisation assay; genetic mapping; gene expression control

KW protein identification; signal transduction pathway;  
KM metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000: 2000EP-0301439.  
XX  
XX 25-FEB-1999: 99US-0121825.  
PR 05-MAR-1999: 99US-0123180.  
PR 09-MAR-1999: 99US-0123548.  
PR 23-MAR-1999: 99US-0125788.  
PR 25-MAR-1999: 99US-0126264.  
PR 29-MAR-1999: 99US-0126785.  
PR 01-APR-1999: 99US-0127462.  
PR 06-APR-1999: 99US-0128234.  
PR 08-APR-1999: 99US-0128714.  
PR 16-APR-1999: 99US-0129845.  
PR 19-APR-1999: 99US-0130077.  
PR 21-APR-1999: 99US-0130449.  
PR 23-APR-1999: 99US-0130510.  
PR 28-APR-1999: 99US-0130891.  
PR 30-APR-1999: 99US-0131449.  
PR 30-APR-1999: 99US-0132048.  
PR 04-MAY-1999: 99US-0132407.  
PR 05-MAY-1999: 99US-0132484.  
PR 06-MAY-1999: 99US-0132485.  
PR 06-MAY-1999: 99US-0132486.  
PR 07-MAY-1999: 99US-0132487.  
PR 11-MAY-1999: 99US-0132863.  
PR 14-MAY-1999: 99US-0134218.  
PR 14-MAY-1999: 99US-0134219.  
PR 14-MAY-1999: 99US-0134221.  
PR 18-MAY-1999: 99US-0134370.  
PR 19-MAY-1999: 99US-0134768.  
PR 19-MAY-1999: 99US-0134941.  
PR 20-MAY-1999: 99US-0135124.  
PR 21-MAY-1999: 99US-0135353.  
PR 24-MAY-1999: 99US-0135629.  
PR 25-MAY-1999: 99US-0136021.  
PR 27-MAY-1999: 99US-0136782.  
PR 28-MAY-1999: 99US-0136789.  
PR 01-JUN-1999: 99US-0137222.  
PR 03-JUN-1999: 99US-0137528.  
PR 04-JUN-1999: 99US-0137502.  
PR 07-JUN-1999: 99US-0137724.  
PR 08-JUN-1999: 99US-0138094.  
PR 10-JUN-1999: 99US-0138540.  
PR 10-JUN-1999: 99US-0138847.  
PR 14-JUN-1999: 99US-0139119.  
PR 16-JUN-1999: 99US-0139452.  
PR 16-JUN-1999: 99US-0139453.  
PR 17-JUN-1999: 99US-0139454.  
PR 18-JUN-1999: 99US-0139454.  
PR 18-JUN-1999: 99US-0139455.  
PR 18-JUN-1999: 99US-0139456.  
PR 18-JUN-1999: 99US-0139457.  
PR 18-JUN-1999: 99US-0139458.  
PR 18-JUN-1999: 99US-0139459.  
PR 18-JUN-1999: 99US-0139460.  
PR 18-JUN-1999: 99US-0139461.  
PR 18-JUN-1999: 99US-0139462.  
PR 18-JUN-1999: 99US-0139463.  
PR 18-JUN-1999: 99US-0139463.  
PR 18-JUN-1999: 99US-0139750.  
PR 18-JUN-1999: 99US-0139763.  
PR 21-JUN-1999: 99US-0139817.  
PR 22-JUN-1999: 99US-0139899.  
PR 23-JUN-1999: 99US-0140353.  
PR 23-JUN-1999: 99US-0140354.  
  
PR 24-JUN-1999: 99US-0140695.  
PR 28-JUN-1999: 99US-0140823.  
PR 29-JUN-1999: 99US-0140991.  
PR 30-JUN-1999: 99US-0141287.  
PR 01-JUL-1999: 99US-0141842.  
PR 02-JUL-1999: 99US-0142154.  
PR 06-JUL-1999: 99US-0142390.  
PR 08-JUL-1999: 99US-0142803.  
PR 09-JUL-1999: 99US-0142920.  
PR 12-JUL-1999: 99US-0142977.  
PR 13-JUL-1999: 99US-0143542.  
PR 14-JUL-1999: 99US-0143624.  
PR 15-JUL-1999: 99US-0144005.  
PR 16-JUL-1999: 99US-0144085.  
PR 16-JUL-1999: 99US-0144086.  
PR 19-JUL-1999: 99US-0144325.  
PR 19-JUL-1999: 99US-0144331.  
PR 19-JUL-1999: 99US-0144332.  
PR 19-JUL-1999: 99US-0144333.  
PR 19-JUL-1999: 99US-0144334.  
PR 19-JUL-1999: 99US-0144335.  
PR 20-JUL-1999: 99US-0144352.  
PR 20-JUL-1999: 99US-0144632.  
PR 20-JUL-1999: 99US-0144684.  
PR 21-JUL-1999: 99US-0144814.  
PR 21-JUL-1999: 99US-0145086.  
PR 22-JUL-1999: 99US-0145088.  
PR 22-JUL-1999: 99US-0145085.  
PR 22-JUL-1999: 99US-0145087.  
PR 22-JUL-1999: 99US-0145089.  
PR 22-JUL-1999: 99US-0145192.  
PR 23-JUL-1999: 99US-0145195.  
PR 23-JUL-1999: 99US-0145218.  
PR 23-JUL-1999: 99US-0145224.  
PR 26-JUL-1999: 99US-0145276.  
PR 27-JUL-1999: 99US-0145913.  
PR 27-JUL-1999: 99US-0145918.  
PR 27-JUL-1999: 99US-0145919.  
PR 28-JUL-1999: 99US-0145951.  
PR 02-AUG-1999: 99US-0146386.  
PR 02-AUG-1999: 99US-0146388.  
PR 02-AUG-1999: 99US-0146389.  
PR 03-AUG-1999: 99US-0147038.  
PR 04-AUG-1999: 99US-0147204.  
PR 05-AUG-1999: 99US-0147302.  
PR 05-AUG-1999: 99US-0147192.  
PR 06-AUG-1999: 99US-0147260.  
PR 06-AUG-1999: 99US-0147303.  
PR 09-AUG-1999: 99US-0147416.  
PR 09-AUG-1999: 99US-0147493.  
PR 09-AUG-1999: 99US-0147935.  
PR 10-AUG-1999: 99US-0148171.  
PR 11-AUG-1999: 99US-0148319.  
PR 12-AUG-1999: 99US-0148341.  
PR 13-AUG-1999: 99US-0148565.  
PR 13-AUG-1999: 99US-0148684.  
PR 16-AUG-1999: 99US-0149368.  
PR 17-AUG-1999: 99US-0149175.  
PR 18-AUG-1999: 99US-0149426.  
PR 20-AUG-1999: 99US-0149722.  
PR 20-AUG-1999: 99US-0149723.  
PR 20-AUG-1999: 99US-0149929.  
PR 23-AUG-1999: 99US-0149902.  
PR 23-AUG-1999: 99US-0149930.  
PR 25-AUG-1999: 99US-0150566.  
PR 26-AUG-1999: 99US-0150884.  
PR 27-AUG-1999: 99US-0151065.  
PR 27-AUG-1999: 99US-0151066.  
PR 27-AUG-1999: 99US-0151080.  
PR 30-AUG-1999: 99US-0151303.  
PR 31-AUG-1999: 99US-0151438.  
PR 01-SEP-1999: 99US-0151930.

```

PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match

Best Local Similarity 9.18; Score 30.6; DB 21; Length 1147;  
Matches 72; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```

OY 90 CCTCTGCGNCTGATGAGGACCACTATGTTGATTCATCAAGCTTACCCGTTGNTACAA 149
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 378 CTTCTACTTCTTCTTCAATCTCATCTTTTTCATTTTGAACATCTCTGATATAGATACC 319
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 150 GTTAGAGCTAAAGAGAAATAGTGGAGTCTTCTGTGAGACCTATCCGATACCCCGNNT 209
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 318 CTAGACTACTCCATGCGATAGCACTCGTGTCTAACCCCGAGCCAGAGAGTCTCCACGT 259
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 210 GGCCCTGAGNTCCAGAGGNGCTGTGTC 238
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 258 CAGCAGCAAACTCTCAACTCCATGTGTC 230
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 4

AAAC40084/c  
ID AAC40084 standard; DNA; 1626 BP.

AC AAC40084;

DF 17-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 26387.

```

KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
PN EPI033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 29-MAR-1999; 99US-0126264.
XX 01-APR-1999; 99US-0126785.
XX 06-APR-1999; 99US-0127462.
XX 08-APR-1999; 99US-0128234.
XX 16-APR-1999; 99US-0128714.
XX 19-APR-1999; 99US-0128845.
XX 21-APR-1999; 99US-0130077.
XX 23-APR-1999; 99US-0130449.
XX 28-APR-1999; 99US-0130510.
XX 30-APR-1999; 99US-0131449.
XX 04-MAY-1999; 99US-0132048.
XX 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 11-MAY-1999; 99US-0132863.
XX 14-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 18-MAY-1999; 99US-0134221.
XX 19-MAY-1999; 99US-0134370.
XX 20-MAY-1999; 99US-0134768.
XX 21-MAY-1999; 99US-0135124.
XX 24-MAY-1999; 99US-0135553.
XX 25-MAY-1999; 99US-0135629.
XX 27-MAY-1999; 99US-0136021.
XX 28-MAY-1999; 99US-0136392.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 14-JUN-1999; 99US-0138847.
XX 16-JUN-1999; 99US-0139119.
XX 17-JUN-1999; 99US-0139453.
XX 18-JUN-1999; 99US-0139452.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139464.
XX 18-JUN-1999; 99US-0139465.
XX 21-JUN-1999; 99US-0139617.
XX 22-JUN-1999; 99US-0139899.
XX 23-JUN-1999; 99US-0140353.

```

```
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141642.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145145.
PR 22-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145228.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145226.
PR 27-JUL-1999; 99US-0145591.
PR 27-JUL-1999; 99US-0145591.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0149368.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.

PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157171.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158322.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 9.18; Score 30.6; DB 21; Length 1626;
Best Local Similarity 48.3%; Pred. No. 2.1;
Matches 72; Conservative 0; Mismatches 77; Indels 0; Caps 0;

QY 90 CTTCTGGGCTGCATGAGGACCACTATGTTGNAATTCATACGTCACCCGTTGNTACAA 149
DB 846 CTTCTACTCTCTTTCATATCATCTTTTCGATTTTGACATCTCGAATAGATACAC 787
QY 150 GTGTAGNCTCAAGAGAAATAGTGGAGTCTTCTGTGAGACCTATGTGATCCCGNNT 209
DB 786 CTAGCTCTCTCATGCGATGACATCTCCCTGTCTTAACCCGCGACGAGAGAGTCTTCACGT 727
QY 210 GGCCCTGAGNTTCCAGAGGCGCTGTCG 238
DB 726 CAGCAGCAAACTCTCAAGCTCATGTCG 698

RESULT 5
AAQ35988
ID AAQ35988 standard; DNA; 5534 BP.
XX
AC AAQ35988;
XX
DT 18-JUN-1993 (first entry)
XX
DE Tomato hsp80 genomic clone Ghsp-4 (locus A).
```

XX	WPI: 1993-075789/09.	
DR	P-PsDB; AAR32549.	
PT	New DNA constructs - comprising tomato heat shock protein 80 gene	
PT	non-coding region and gene region of interest, useful for	
PT	producing transgenic plants with modified phenotypes)	
PS	Disclosure; Fig 2; 33pp; English.	
XX		
CC	The sequence given represents the 5' non-coding region, all of the	
CC	coding region and the 3' non-coding region of the tomato heat shock	
CC	protein 80.5 (hsp80) genomic clone Ghsp-41locus A). Tomato hsp80 is	
CC	characterised by having a mRNA of about 2.3 kb, two introns, a	
CC	predicted pl of about 4.69 and a molecular weight of 80,479.8 daltons.	
CC	The polypeptides encoded by tomato hsp80 have substantial homology	
CC	with the polypeptides encoded by related heat shock proteins from a	
CC	variety of prokaryotic and eukaryotic species. There are two hsp80	
CC	genes in tomato, identified as locus A and locus B. The low copy	
CC	number of the hsp80 gene is a useful indicator of the strength of the	
CC	promoter. Tomato hsp80 message is found in abundance in most, if not	
CC	all meristematic tissue eg, flower meristems, early fruit, and root	
CC	and shoot apices. hsp80 message is detected in floral primordia of	
CC	immature fruit, where it increases until the fruit begins to swell.	
CC	The mRNA levels observed in normal mature leaves is approx. 1/10th of	
CC	the level observed in meristematic tissue.	
XX		
SQ	Sequence 5534 BP; 1793 A; 810 C; 1066 G; 1865 T; 0 other;	
	Query Match	9.0%; Score 30.4; DB 14; Length 5534;
	Best Local Similarity	55.6%; Pred. No. 4.1;
	Matches	55; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
OY	104 TGAGGCACCATATGTTGATTCATCTACCTACCCCGTTGNTCAAGTATGNCITCAAG	163
DB	255 tgaagccaaagcatctctaaagatcttggtcaccgcgatgtccaatgacatgagaag	314
OY	164 GAAGATAGCTGGAGCTTCTGTGTGACACTATCTGAATC	202
DB	315 taaataataatgactgtgtgtgttagtaataatgattc	353
RESULT	6	
AAAX20446		
ID	AAAX20446 standard; DNA: 965 BP.	
.XX		
AC	AAAX20446;	
XX		
DT	04-MAY-1999 (first entry)	
XX		
DE	Human secreted protein gene 35.	
XX		
Human; secreted protein; fusion protein; gene therapy; protein therapy;		
diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;		
developmental abnormality; foetal deficiency; brain; hepatic; lymphoma;		
immune system; asthma; lymphocytic disease; lymphoma;		
inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;		
cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;		
osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;		
endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.		
XX		
OS	Homo sapiens.	
XX		
PN	W09906423-A1.	
XX		
PD	11-FEB-1999.	
XX		
PF	29-JUL-1998; 98WO-US15949.	
XX		
PR	19-AUG-1997; 97US-0056730.	
PR	30-JUL-1997; 97US-0054209.	
PR	30-JUL-1997; 97US-0054211.	



AC AAC93406;  
AD  
DE 26-FEB-2001 (first entry)  
XX  
DE Human secreted protein gene 43 SEQ ID NO:53.  
XX  
KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;  
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;  
KW cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;  
KW fungicide; ophthalmological; gene therapy; autoimmune disease; infection  
KW hyperproliferative disorder; cardiovascular disorder; angiogenesis;  
KW cerebrovascular disorder; nervous system disorder; ocular disorder;  
KW wound healing; skin aging; food additive; preservative; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200061620-A1.  
PD  
PD 19-OCT-2000.  
XX  
XX 06-APR-2000; 2000WO-US09069.  
PA  
PA 09-APR-1999; 9GUS-0128702.  
PR 20-JAN-2000; 2000US-0177049.  
PP  
PP (HUMA-) HUMAN GENOME SCI INC.  
PA (ROSE/) ROSEN C A.  
PI  
PI Rosen CA, Ruben SM, Komatsoulis G;  
XX  
DR MPI: 2000-619225/59.  
DR P-PSDB; AAB51662.  
PT  
PT Isolated nucleic acid molecule encoding a human secreted protein is  
PT used in preventing, treating or ameliorating a medical condition -  
XX  
Claim 1; Page 458; 540pp; English.

The polynucleotide sequences given in AAC93364 to AAC93412 encode the human secreted proteins given in AAB51620 to AAB51668. AAB51669 to AAB51722 represent human secreted polypeptide sequences and proteins homologous to them, which are given in the exemplification of the present invention. Human secreted proteins have activities based on the tissues and cells the genes are expressed in. Examples of activities include: antihypertensive; immunosuppressive; antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide; fungicide; and ophthalmological. The polynucleotides and polypeptides can be used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Disorders which are diagnosed or treated include autoimmune diseases, hyperproliferative disorders, cardiovascular disorders, cerebrovascular disorders, angiodysplasia, nervous system disorders, infections caused by bacteria, viruses and fungi and ocular disorders. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. AAC93355 to AAC93363 and AAB51619 represent sequences which are used in the exemplification of the present invention.

Sequence 1278 BP; 259 A; 355 C; 379 G; 278 T; 7 other:

Query Match 9.0%; Score 30.2; DB 21; Length 1278;  
Best Local Similarity 57.5%; Pred. No. 2.5;  
Matches 50; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

173 TGGAGCTCTTCGTAGACTATCGATCCGCCGNNNTGGCCCTTGAGNMTTCCAGAGGNC 232  
||||| ||||| | ||| ||||| | |||||

[illegible]









CC The invention involves production of transgenic plants containing  
CC DNA comprising geminivirus AC1 or CI wild-type or mutant sequences  
CC that negatively interfere in trans with geminiviral replication  
CC during infection. Such transgenic plants are resistant to viral  
CC infection. The AC1/C1 genes are especially from ToMoV, tomato  
CC yellow leaf curl virus or bean golden mosaic geminivirus (see  
CC AM933282-93) and encode polypeptides (see AA034324-35) that have  
CC mutations in the highly conserved DNA-nicking and/or the NTP-binding  
CC domains.  
XX  
Sequence 1169 BP; 363 A; 281 C; 255 G; 270 T; 0 other:

Sequence 1169 BP; 363 A; 281 C; 255 G; 270 T; 0 other;

Query Match	8.8%	Score 29.6;	DB 18;	Length 1169;
Best Local Similarity	63.1%;	Pred. No. 3.9;		
Matches 41;	Conservative 0;	Mismatches 24;	Indels 0;	Gaps 0;

205 CGNNITGGCCCTGAGNTTCCAGAGGGNCTGTGCTGCCCATCGCCCTAGCAGGGTTCAAGNA 264  
 || || ||| ||||| |||| | ||||| || ||||| |  
 660 CGGTTCAAGCTCTCGCGCCAGAGAGAACTGTAAATATCATCTGTCGAGGTGTATTCAGAA 719

cy	265	AAGCG	269
Db	720	caggg	724

Search completed: August 24, 2001, 03:53:37  
Job time: 1418 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2001, 03:12:34 ; Search time 57.67 Seconds  
(without alignments)  
1102.974 Million cell updates/sec

Title: US-08-224-621-74

Perfect score: 336  
1 CCGTCGACATCTTTTCTATGC.....CCCCNTGCGTCCCAACCA 336

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_MA:\*

- 1: /cgn2\_6/ptodata/2/1na/5A.COMB.seq:\*
- 2: /cgn2\_6/ptodata/2/1na/5B.COMB.seq:\*
- 3: /cgn2\_6/ptodata/2/1na/6A.COMB.seq:\*
- 4: /cgn2\_6/ptodata/2/1na/6B.COMB.seq:\*
- 5: /cgn2\_6/ptodata/2/1na/PCTUS.COMB.seq:\*
- 6: /cgn2\_6/ptodata/2/1na/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	28.4	8.5	7286	5 PCT-US95-11684-1	Sequence 1, Appli
C 2	28.2	8.4	3416	2 US-08-357-642A-2	Sequence 2, Appli
C 3	28.2	8.4	3416	2 US-08-460-626-2	Sequence 2, Appli
C 4	27.8	8.3	49136	4 US-09-422-869-1	Sequence 1, Appli
C 5	27.2	8.1	940	2 US-08-713-000-3	Sequence 3, Appli
C 6	27.2	8.1	940	2 US-08-975-316-3	Sequence 3, Appli
C 7	27.2	8.1	940	4 US-09-211-710-3	Sequence 3, Appli
C 8	27.2	8.1	1512	3 US-09-911-853-6	Sequence 6, Appli
C 9	27.2	8.1	1512	4 US-09-479-409-6	Sequence 6, Appli
C 10	27.2	8.1	1785	2 US-08-975-316-48	Sequence 4, Appli
C 11	27.2	8.1	17612	3 US-08-911-853-29	Sequence 29, Appli
C 12	27.2	8.1	17612	3 US-09-479-409-29	Sequence 29, Appli
C 13	27.2	8.0	17949	4 US-09-087-465-3	Sequence 3, Appli
C 14	26.4	7.9	1329	1 US-08-278-630A-8	Sequence 8, Appli
C 15	26.4	7.9	2890	3 US-08-848-810-1	Sequence 1, Appli
C 16	26.4	7.9	3150	4 US-08-943-768-1	Sequence 1, Appli
C 17	26.4	7.9	72928	3 US-09-009-913-1	Sequence 1, Appli
C 18	26.2	7.8	1209	1 US-08-314-309A-5	Sequence 5, Appli
C 19	26.2	7.8	1513	1 US-08-314-309A-2	Sequence 2, Appli
C 20	26.2	7.8	1546	1 US-08-314-309A-3	Sequence 3, Appli
C 21	26.2	7.8	1952	1 US-08-333-358-1	Sequence 1, Appli
C 22	26.2	7.8	1952	1 US-08-463-694-1	Sequence 1, Appli
C 23	26.2	7.8	1952	1 US-08-694-501-1	Sequence 1, Appli
C 24	26.2	7.8	3172	1 US-08-314-309A-1	Sequence 1, Appli
C 25	26.2	7.7	2200	2 US-08-819-825-1	Sequence 1, Appli
C 26	26.2	7.7	2200	2 US-09-163-642-1	Sequence 1, Appli
C 27	25.8	7.7	1356	2 US-08-484-126-4	Sequence 4, Appli

C 28	25.8	7.7	3449	3 US-09-049-672A-16	Sequence 16, Appli
C 29	25.6	7.6	1088	4 US-09-077-675A-6	Sequence 6, Appli
C 30	25.6	7.6	1122	4 US-09-077-675A-9	Sequence 9, Appli
C 31	25.6	7.6	1596	4 US-08-416-544B-8	Sequence 8, Appli
C 32	25.6	7.6	1875	4 US-09-422-869-21	Sequence 21, Appli
C 33	25.6	7.6	2115	4 US-08-796-101-48	Sequence 48, Appli
C 34	25.6	7.6	2303	2 US-08-480-229C-9	Sequence 9, Appli
C 35	25.6	7.6	2303	2 US-08-659-235C-9	Sequence 9, Appli
C 36	25.6	7.6	2308	2 US-08-480-229C-28	Sequence 28, Appli
C 37	25.6	7.6	2308	2 US-08-659-235C-28	Sequence 28, Appli
C 38	25.6	7.6	2421	3 US-08-851-843A-51	Sequence 51, Appli
C 39	25.6	7.6	2421	4 US-08-974-549A-218	Sequence 218, App
C 40	25.6	7.6	5275	4 US-08-796-101-49	Sequence 49, Appli
C 41	25.6	7.6	49136	4 US-09-422-869-1	Sequence 1, Appli
C 42	25.4	7.6	2060	1 US-08-480-547A-11	Sequence 11, Appli
C 43	25.4	7.6	2060	1 US-08-250-847B-11	Sequence 11, Appli
C 44	25.4	7.6	2060	2 US-08-463-949A-11	Sequence 11, Appli
C 45	25.4	7.6	2060	3 US-08-464-410A-11	Sequence 11, Appli

## ALIGNMENTS

RESULT 1  
PCT-US95-11684-1/c  
; Sequence 1, Application PC/TUS9511684  
; GENERAL INFORMATION:  
; APPLICANT: THE SCRIPPS RESEARCH INSTITUTE  
; TITLE OF INVENTION: CYTOTACTIN DERIVATIVES THAT STIMULATE  
; TITLE OF INVENTION: ATTACHMENT AND NEURITE OUTGROWTH, AND METHODS OF MAKING  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: The Scripps Research Institute, Office of  
; ADDRESSEE: Patent Counsel  
; STREET: 10666 North Torrey Pines Road, TPC 8  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/11684  
; FILING DATE: 14-SEP-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/308,359  
; FILING DATE: 16-SEP-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Logan, April C.  
; REGISTRATION NUMBER: 33,950  
; REFERENCE/DOCKET NUMBER: BEC0019P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-554-2937  
; TELEFAX: 619-554-6312  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7286 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 55..6654  
; OTHER INFORMATION: /product= "cytotactin"  
PCT-US95-11684-1



APPLICANT: HORIKAWA, YUKIO  
APPLICANT: ODA, NAOKISA  
APPLICANT: COX, NANCY J.  
APPLICANT: SREENAN, SEAMUS  
APPLICANT: ZHOU, YUN-PING  
APPLICANT: OTANI, KENICHI  
APPLICANT: HANIS, CRAIG L.  
APPLICANT: BELL, GRAEME I.  
TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES  
FILE REFERENCE: ARCD:307  
CURRENT APPLICATION NUMBER: US/09/422,869  
EARLIER FILING DATE: 1989-10-21  
EARLIER APPLICATION NUMBER: 60/134,175  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 49136  
TYPE: DNA  
ORGANISM: Human  
-09-422-869-1

Query Match 8.38; Score 27.8; DB 4; Length 49136;  
Best Local Similarity 53.88; Pred. No. 15;  
Matches 50; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

OY 189 GACCTATCTGAATCCCGGNNNTGCCCTGAGNTCCAGAGGNCCTGGTCGCCATGCC 248  
DB 35030 ggcctccagagatccctccagaccctggagccctgctccctcgccgcgagctg 35089  
OY 249 TACGAGGCTTCAGNMAAGGCGCCGNCATGG 281  
DB 35090 gtgcagctctctgcgaaggccctgactgtg 35122

RESULT 5  
US-08-713-000-3/c  
Sequence 3, Application US/08713000  
Patent No. 5850020  
GENERAL INFORMATION:  
APPLICANT: Bloksberg, Leonard N.  
APPLICANT: Havukkala, Ilkka  
APPLICANT: Grierson, Alastair  
TITLE OF INVENTION: MATERIALS AND METHODS FOR THE  
MODIFICATION OF PLANT LIGNIN CONTENT  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Speckman Picard PLLC  
STREET: 2601 Elliott Avenue, Suite 4185  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/713,000  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Sleath, Janet  
REGISTRATION NUMBER: 37,007  
REFERENCE/DOCKET NUMBER: 11000.1003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-269-0563  
TELEFAX: 206-269-0563  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 940 base pairs

TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-713-000-3

Query Match 8.18; Score 27.2; DB 2; Length 940;  
Best Local Similarity 50.54; Pred. No. 2.7;  
Matches 53; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

OY 232 CTGGTCCATCCATCCGCTAGAGGCTTCAGNMAAGGCGCCGNCATGGCAGTCCCTGG 291  
DB 196 CTGGTCCGTTATCTGCACGCTGCTCCAGACAGGCTCCTGCGGACCTTGC 137  
OY 292 NCAGNMAANGANTTGNCCCAACCCNTTGGTGTCCCAACCA 336  
DB 136 TCTGATGTCTGTGTGTTACACAGCTCCGCTATTCCTCCATCCA 92

RESULT 6  
US-08-975-316-3/c  
Sequence 3, Application US/08975316  
Patent No. 5952486  
GENERAL INFORMATION:  
APPLICANT: Bloksberg, Leonard N., HAVUKKALA, Ilkka  
APPLICANT: and GRIERSON, Alastair W.  
TITLE OF INVENTION: MATERIALS AND METHODS FOR  
THE MODIFICATION OF PLANT LIGNIN CONTENT  
NUMBER OF SEQUENCES: 88  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Offices of Ann W. Speckman  
STREET: 2601 Elliott Avenue, Suite 4185  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/975,316  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/713,000  
FILING DATE: September 11, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: SLEATH, Janet  
REGISTRATION NUMBER: 37,007  
REFERENCE/DOCKET NUMBER: 11000/1003c1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-269-0563  
TELEFAX: 206-269-0563  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 940 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-975-316-3

Query Match 8.18; Score 27.2; DB 2; Length 940;  
Best Local Similarity 50.54; Pred. No. 2.7;  
Matches 53; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

OY 232 CTGGTCCATCCATCCGCTAGAGGCTTCAGNMAAGGCGCCGNCATGGCAGTCCCTGG 291  
DB 196 CTGGTCCGTTATCTGCACGCTGCTCCAGACAGGCTCCTGCGGACCTTGC 137

QY 292 NCAGNANGANGANTTGGNCCCAACCCNTTGGTTCCCAACCA 336  
DB 136 TGTGAATGTCCTGGTGTTCACACAGCTCCGCTATTCCTCCATTC 92

RESULT 7  
US-09-211-710-3/c

Sequence 3, Application US/09211710A  
Patent No. 6204434  
GENERAL INFORMATION:  
APPLICANT: Bloksberg, Leonard N.  
APPLICANT: Havukkala, Ilkka  
TITLE OF INVENTION: Grierson, Alastair  
TITLE OF INVENTION: Materials and Methods for the  
FILE REFERENCE: 11000.1003c3  
CURRENT APPLICATION NUMBER: US/09/211.710A  
CURRENT FILING DATE: 1998-12-14  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 3  
LENGTH: 940  
TYPE: DNA  
ORGANISM: Pinus radiata  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (463)...(463)  
US-09-211-710-3

Query Match  
Best Local Similarity 8.1%; Score 27.2; DB 4; Length 940;  
Matches 53; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 232 CTGTCCTCCATCGCCATCAAGNNAAGGGCCGCGCATGCTCTTG 231  
DB 196 CTGTCCTCCATCGCCATCAAGNNAAGGGCCGCGCATGCTCTTG 137  
QY 292 NCAGNANGANGANTTGGNCCCAACCCNTTGGTTCCCAACCA 336  
DB 136 TGTGAATGTCCTGGTGTTCACACAGCTCCGCTATTCCTCCATTC 92

RESULT 8

US-08-911-853-6  
Sequence 6, Application US/08911853  
Patent No. 6046710  
GENERAL INFORMATION:

APPLICANT: Gerritse, Gijbert  
APPLICANT: Quax, Wilhelmus J.  
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESS: Genencor International  
STREET: 925 Page Mill Road  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1013  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/911.853  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/699,092  
FILING DATE: 16-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Glaister, Debra J

REGISTRATION NUMBER: 33,888  
REFERENCE/DOCKET NUMBER: GC361-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-846-7620  
TELEFAX: 650-845-6504  
INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1512 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-911-853-6

Query Match  
Best Local Similarity 8.1%; Score 27.2; DB 3; Length 1512;  
Matches 41; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 173 TGGAGCTCTTCTGTGAGACCTATGATCCCGGNTTGGCTGAGNTTCCAGAGGNC 232  
DB 896 TGGAGGATCGGCGGAGCCAGACCGGATGACATGACCTCGCCGCGGCC 955  
QY 233 TGGTCGTC 240  
DB 956 TGGCGGCC 963

RESULT 9

US-09-479-409-6  
Sequence 6, Application US/09479409  
Patent No. 6225106  
GENERAL INFORMATION:

APPLICANT: Gerritse, Gijbert  
APPLICANT: Quax, Wilhelmus J.  
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESS: Genencor International  
STREET: 925 Page Mill Road  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1013  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/479,409  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/911,853  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Glaister, Debra J  
REGISTRATION NUMBER: 33,888  
REFERENCE/DOCKET NUMBER: GC361-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-846-7620  
TELEFAX: 650-845-6504  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1512 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-479-409-6

Query Match  
Best Local Similarity 8.1%; Score 27.2; DB 4; Length 1512;  
Matches 60.3%; Pred. No. 3.5;





Query Match 7.9%; Score 26.4; DB 1; Length 1329;  
Best Local Similarity 49.2%; Pred. No. 6.5;  
Matches 63; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

Db 180 CTCCCTCTGGGAGAGAGACACCTTCCCTAAACATATTTTGACTGTGTGATTT 239  
OY 82 ACTCGAGACCCCTGTGCATGAGCAGCACCATATGTTGATATCATCAGTACCCGT 141  
Db 240 AGTGGAGACCATTTGGATGACCCAGAACCCGATTTTGGAGATGTTGCCGCTCTCCGG 299  
OY 142 TGNATCAA 149  
Db 300 GGGAAGAA 307

## RESULT 15

US-08-848-810-1  
: Sequence 1, Application US/08848810  
: Patent No. 6074851  
: GENERAL INFORMATION:  
: APPLICANT: Deidel Jr., M. R.  
: APPLICANT: Yem, A. W.  
: APPLICANT: Wilson, C. L.  
: TITLE OF INVENTION: Catalytic Macro Molecules Having DCD25B  
: NUMBER OF SEQUENCES: 45  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Pharmacia & Upjohn Company  
: STREET: 301 Henrietta Street  
: CITY: Kalamazoo  
: STATE: MI  
: COUNTRY: USA  
: ZIP: 49001  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patent Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/848,810  
: FILING DATE:  
: CLASSIFICATION: 435  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Wootton, Thomas A.  
: REGISTRATION NUMBER: 35,004  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 616-833-7914  
: TELEFAX: 616-833-6897  
: INFORMATION FOR SEQ ID NO: 1:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 2890 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: CDNA  
: HYPOTHEICAL: NO  
: ANTI-SENSE: NO  
: US-08-848-810-1

Query Match 7.9%; Score 26.4; DB 3; Length 2890;

Best Local Similarity 48.6%; Pred. No. 10;

Matches 54; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

OY 216 GAGNTTCCAGAGGNCCTGTCCTCCATCCGCTAGCAGGGTTCAGNAAGGGGCCCGC 275  
Db 578 GAGCTGCCAGCAGCTCTGGGGAATCCAGCTCCACCCATGCTCTGGCAGAGTGGGCCAGC 637  
OY 276 NCATGGCAGCTCTTGNCAGNAGNANGANTTGGNCCCAACCCCTTGGT 326  
Db 638 CGCAGGGGAAGCCTTGGCCAGAGACCCAGCTCGGCCGCCGACCTGATGTGT 688

Search completed: August 24, 2001, 03:51:27  
Job time: 2333 sec

Fri Aug 24 10:01:25 2001

us-08-224-621-74.rni

---



```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*

```

```

190: gb_est110:*
191: gb_est111:*
192: gb_hic:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_inv4:*
207: em_gss_inv5:*
208: em_gss_inv6:*
209: em_gss_inv7:*
210: em_gss_inv8:*
211: em_gss_inv9:*
212: em_gss_inv10:*
213: em_gss_inv11:*
214: em_gss_inv12:*
215: em_gss_inv13:*
216: em_gss_inv14:*
217: em_gss_inv15:*
218: em_gss_inv16:*
219: em_gss_inv17:*
220: em_gss_inv18:*
221: em_gss_inv19:*
222: em_gss_inv20:*
223: em_gss_inv21:*
224: em_gss_inv22:*
225: em_gss_inv23:*
226: em_gss_inv24:*
227: em_gss_inv25:*
228: em_gss_inv26:*
229: em_gss_inv27:*
230: em_gss_inv28:*
231: em_gss_inv29:*
232: em_gss_inv30:*
233: em_gss_inv31:*
234: em_gss_inv32:*
235: em_gss_inv33:*
236: em_gss_inv34:*
237: em_gss_inv35:*
238: em_gss_inv36:*
239: em_gss_inv37:*
240: em_gss_inv38:*
241: em_gss_inv39:*
242: em_gss_inv40:*
243: em_gss_inv41:*
244: em_gss_inv42:*
245: em_gss_inv43:*
246: em_gss_inv44:*
247: em_gss_inv45:*
248: em_gss_inv46:*
249: em_gss_inv47:*
250: em_gss_inv48:*
251: em_gss_inv49:*
252: em_gss_inv50:*
253: em_gss_inv51:*
254: em_gss_inv52:*
255: em_gss_inv53:*
256: em_gss_inv54:*
257: em_gss_inv55:*
258: em_gss_inv56:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



```

/clone="E1BR036H10"
/clone_lib="Bos taurus brain fetus"
/tissue_type="brain"
/dev_stage="fetus"
/lab_host="DH10B"
/Note="Vector: pZL1; Site_1: SalI; Site_2: NotI; Poly A
was deleted from a NotI site"
BASE COUNT      157 a      133 c      123 g      126 t
ORIGIN

```

```

Query Match      24.4% Score 82; DB 32; Length 539;
Best Local Similarity 89.4%; Pred. No. 4.3e-15;
Matches 110; Conservative 0; Mismatches 11; Indels 2; Gaps 2;

```

```

OY 2 GCTGCATCTTTTCTATGCTCTCCCTGCTGGCGGTGAGATACAGACGAAACG 61
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 384 GCTGCATCTTTTCTATGCTCTCCCTGCTGGCGGTGAGATACAGACGAAACG 443
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
62 GACAGCTCGTTCATGATGACGACCTGCTGCGTGCATGAGCACCACCTATGTTG 121
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
444 GAGAGCTCAT-TCATGATGAGACTGCGATCTCAACG-CTGCATGAGCACCACCTATGTTG 501
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 122 NAT 124
    |
DB 502 ATT 504
    |

```

```

RESULT 3
LOCUS AV663814 539 bp mRNA EST 25-AUG-2000
DEFINITION AV663814 Bos taurus brain fetus Bos taurus cDNA clone E1BR038G07
VERSION AV663814 GI:9922844
KEYWORDS EST.
SOURCE Bos taurus
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 539)
Sugimoto,Y., Hirotsune,S., Takasuga,A., Itoh,R., Jitohzono,A. and
Suzuki,H.
bovine cDNA sequencing
Unpublished (2000)
Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazusugi@cocoa.ocn.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.

```

```

FEATURES
Source
1..539
Location/Qualifiers
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone="E1BR038G07"
/clone_lib="Bos taurus brain fetus"
/tissue_type="brain"
/dev_stage="fetus"
/lab_host="DH10B"
/Note="Vector: pZL1; Site_1: SalI; Site_2: NotI; Poly A
was deleted from a NotI site"
BASE COUNT      138 a      144 c      129 g      128 t
ORIGIN

```

```

Query Match      24.4% Score 82; DB 32; Length 539;
Best Local Similarity 89.4%; Pred. No. 4.3e-15;
Matches 110; Conservative 0; Mismatches 11; Indels 2; Gaps 2;

```

```

OY 2 GCTGCATCTTTTCTATGCTCTCCCTGCTGGCGGTGAGATACAGACGAAACG 61
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 310 GCTGCATCTTTTCTATGCTCTCCCTGCTGGCGGTGAGATACAGACGAAACG 369
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 62 GACAGCTCGTTCATGATGACGACCTGCTGCGTGCATGAGCACCACCTATGTTG 121
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 370 GAGAGCTCAT-TCATGATGAGACTGCGATCTCAACG-CTGCATGAGCACCACCTATGTTG 427
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 122 NAT 124
    |
DB 428 ATT 430
    |

```

```

RESULT 4
LOCUS AUI35898 796 bp mRNA EST 24-OCT-2000
DEFINITION AUI35898 PLACEL Homo sapiens cDNA clone PLACEL1003129 5', mRNA
sequence.
AUI35898
VERSION AUI35898.1 GI:10996437
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 796)
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,Y., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
HRI human cDNA project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

```

```

FEATURES
Source
1..796
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PLACEL1003129"
/clone_lib="PLACEL1"
/tissue_type="placenta"
/Note="Vector: pME18SFL3"
BASE COUNT      205 a      207 c      197 g      180 t      7 others
ORIGIN

```

```

Query Match      23.5% Score 78.8; DB 108; Length 796;
Best Local Similarity 87.8%; Pred. No. 4.8e-14;
Matches 108; Conservative 0; Mismatches 13; Indels 2; Gaps 2;

```

```

OY 2 GCTGCATCTTTTCTATGCTCTCCCTGCTGGCGGTGAGATACAGACGAAACG 61
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 313 GCTGCATCTTTTCTATGCTCTCCCTGCTGGCGGTGAGATACAGACGAAACG 372
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 62 GACAGCTCGTTCATGATGACGACCTGCTGCGTGCATGAGCACCACCTATGTTG 121
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 373 GACAGCTCAT-TCATGATGAGACTGCGACCTCGACG-CTGCATGAGCACCACCTATGTTG 430
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 122 NAT 124
    |
DB 431 ATT 433
    |

```

```

RESULT 5

```



AL537924 952 bp mRNA EST 13-FEB-2001  
 LOCUS AL537924 LTI\_FL013.FBrl Homo sapiens cDNA clone CS0DF028Y018 5  
 DEFINITION prime, mRNA sequence.  
 ACCESSION AL537924 GI:12801417  
 VERSION AL537924.1  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 952)  
 AUTHORS Li, W. B., Gruber, C., Jesse, J., and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
 Location/Qualifiers  
 1. 952  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="CS0DF028Y018"  
 /dev\_stage="pooled tissue from post conception fetuses (20  
 week, 24 week and 26 week)"  
 /lab\_host="DH10B"  
 /note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand  
 cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-stranded cDNA was digested with Not I  
 and cloned into the Not I and Eco RV sites of the  
 pCMVSPORT 6 vector. Library was constructed by life  
 Technologies. Contact : Feng Liang Life Technologies, a  
 division of Invitrogen 9800 Medical Center Drive Rockville  
 , Maryland 20850, USA Fax : (1) 301 610 8371 Email :  
 filiang@lifetech.com URL :  
 http://fulllength.invitrogen.com"

BASE COUNT 229 a 224 g 245 t 3 others  
 ORIGIN

Query Match 23.3%; Score 78.4; DB 106; Length 952;  
 Best Local Similarity 87.0%; Pred. No. 6; Be-14;  
 Matches 107; Conservative 1; Mismatches 13; Indels 2; Gaps 2;

QY 2 GCTGATCTTTTCTATGCTCTCCCTGCTGCGCTGATGGAGATACAGACCAAAACG 61  
 |||||  
 290 GCTGATCTTTTCTATGCTCTCCCTGCTGCGCTGATGGAGATACAGACCAAAACG 349  
 |||||

DB 350 GACAGCTCAT-TCAATATGAGACTCGAGACCTCGACG-CTGCATGAGGACCACTATGTGG 407  
 |||||

QY 122 NAT 124  
 Db 408 ATT 410

RESULT 6  
 AL541183 900 bp mRNA EST 16-FEB-2001  
 LOCUS AL541183 LTI\_FL002.PL1 Homo sapiens cDNA clone CS0DE005YD14 5 prime  
 DEFINITION mRNA sequence.  
 ACCESSION AL541183  
 VERSION AL541183.1 GI:12872004  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 900)  
 AUTHORS Li, W. B., Gruber, C., Jesse, J., and Polayes, D.

TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
 Location/Qualifiers  
 1. 900  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="CS0DE005YD14"  
 /clone\_1lb="LTI\_FL002.PL1"  
 /lab\_host="DH10B"  
 /note="Organ: placenta; Vector: pCMVSPORT 6; 1st strand  
 cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-stranded cDNA was digested with Not I  
 and cloned into the Not I and Eco RV sites of the  
 pCMVSPORT 6 vector. Library was constructed by life  
 Technologies. Contact : Feng Liang Life Technologies, a  
 division of Invitrogen 9800 Medical Center Drive Rockville  
 , Maryland 20850, USA Fax : (1) 301 610 8371 Email :  
 filiang@lifetech.com URL :  
 http://fulllength.invitrogen.com"

BASE COUNT 225 a 238 c 213 g 222 t 2 others  
 ORIGIN

Query Match 23.2%; Score 78; DB 106; Length 900;  
 Best Local Similarity 89.8%; Pred. No. 8; Be-14;  
 Matches 106; Conservative 0; Mismatches 10; Indels 2; Gaps 2;

QY 2 GCTGATCTTTTCTATGCTCTCCCTGCTGCGCTGATGGAGATACAGACCAAAACG 61  
 |||||

DB 293 GCTGATCTTTTCTATGCTCTCCCTGCTGCGCTGATGGAGATACAGACCAAAACG 352  
 |||||

QY 62 GACAGCTCAT-TCAATATGAGACTCGAGACCTCGACG-CTGCATGAGGACCACTATGT 119  
 |||||

DB 353 GACAGCTCAT-TCAATATGAGACTCGAGACCTCGACG-CTGCATGAGGACCACTATGT 408  
 |||||

RESULT 7  
 BE032856 547 bp mRNA EST 09-JUL-2000  
 LOCUS BE032856 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.  
 DEFINITION BE032856  
 ACCESSION BE032856  
 VERSION BE032856.1 GI:8327865  
 KEYWORDS EST.  
 SOURCE pig.  
 ORGANISM Sus scrofa  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.  
 REFERENCE 1 (bases 1 to 547)  
 AUTHORS Fahrenkrug, S. C., Treking, B. A., Rohrer, G. A., Smith, T. P. L., Casas, E.,  
 Stone, R. T., Heaton, M. P., Grosse, W. M., Bennett, G. A., Laegreid, W. W.,  
 and Keeler, J. W.  
 TITLE Design and use of two pooled tissue normalized cDNA libraries for  
 EST discovery in swine  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Smith TPL  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smith@email.marc.usda.gov  
 Single pass sequencing. Bases called and alt. trimmed with phred  
 v0.980904.e. Vector identified by cross\_match with the -minscore 18  
 and -minmatch 12 options.  
 PCR primers  
 FORWARD: AGCAACAGCTATGACCAT  
 BACKWARD: GTTTTCCTCAGTCACGACG  
 Plate: 66 row: K column: 23  
 Seq primer: ATTTAGTGACACATATAG.  
 Location/Qualifiers

## source

1. .547  
/organism="Sus scrofa"  
/db\_xref="taxon:9823"  
/clone\_lib="MARC 1pRG"  
/tissue\_type="pooled"  
/lab\_host="DH10B"

/note="Vector: PCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;  
Library made from pooled tissue from day 11, 13, 15, 20,  
and 30 embryos."

BASE COUNT 110 a 178 c 149 g 110 t  
ORIGIN

Query Match 22.4%; Score 75.4; DB 162; Length 547;  
Best Local Similarity 88.1%; Pred. No. 5.1e-13;  
Matches 104; Conservative 0; Mismatches 12; Indels 2; Gaps 2;

QY 7 ATCTTTTCTATGCTCTCCCTGCGCGGTGATGAGATACAGACGACAAACGACAG 66  
|||||  
2 ATCTTTTCTATGCTCTCCCTGCGCGGTGATGAGATACAGACGACAAACGACAG 61  
|||||  
57 CTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 124  
|||||  
Db 62 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 117  
|||||

RESULT 8  
LOCUS BE121364/c 435 bp mRNA EST 13-JUN-2000  
DEFINITION UI-R-CAO-bax-a-08-0-UI.s1 UI-R-CAO Rattus norvegicus cDNA clone  
UI-R-CAO-bax-a-08-0-UI 3', mRNA sequence.  
ACCESSION BE121364  
VERSION BE121364.1 GI:8513469  
KEYWORDS EST.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1 (bases 1 to 435)  
AUTHORS Ronaldo, M.F., Lennon, G. and Soares, M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
COMMENT Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu

Oligo-dt track not found. Not a site shown in beginning of sequence  
is likely internal to the message. cDNA Library Preparation: M.B.  
Soares Lab Clone distribution: clones will be available through  
Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA-NO.

## FEATURES

## source

Location/Qualifiers  
1. .435  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone\_lib="UI-R-CAO-bax-a-08-0-UI"  
/clone\_lib="UI-R-CAO"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pRT73D-Pac (Pharmacia) with a modified  
polylinker. Site\_1: Not I; Site\_2: Eco RI; The UI-R-CAO  
library is a subtracted library derived from the following  
tissues: thalamus, cerebellum, hypothalamus, medulla, pons  
, midbrain, cerebral cortex, corpus striatum, testis, and  
hippocampus. For a detailed description of the library  
from which this clone was derived, please visit our web

BASE COUNT 100 a 106 c 114 g 115 t  
ORIGIN

Query Match 18.6%; Score 62.6; DB 163; Length 435;  
Best Local Similarity 80.5%; Pred. No. 5.2e-09;  
Matches 99; Conservative 0; Mismatches 16; Indels 8; Gaps 2;

QY 2 GCTGATCTTTTCTATGCTCTCCCTGCGCGGTGATGAGATACAGACGACAAACG 61  
|||||  
Db 166 GCTGATCTTTTCTATGCTCTCCCTGCGCGGTGATGAGATACAGACGACAAACG 107  
|||||  
QY 62 GACAGCTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 121  
|||||  
Db 106 GACAGCTGCT-TCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 55  
|||||  
QY 122 NAT 124  
1  
Db 54 ATT 52

RESULT 9  
LOCUS BF400681/c 437 bp mRNA EST 28-NOV-2000  
DEFINITION UI-R-CAO-bhe-c-02-0-UI.s1 UI-R-CAO Rattus norvegicus cDNA clone  
UI-R-CAO-bhe-c-02-0-UI 3', mRNA sequence.  
ACCESSION BF400681  
VERSION BF400681.1 GI:11388656  
KEYWORDS EST.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1 (bases 1 to 437)  
AUTHORS Ronaldo, M.F., Lennon, G. and Soares, M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
COMMENT Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu

Oligo-dt track not found. Not a site shown in beginning of sequence  
is likely internal to the message. cDNA Library Preparation: M.B.  
Soares Lab Clone distribution: clones will be available through  
Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA-NO.

## FEATURES

## source

Location/Qualifiers  
1. .437  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone\_lib="UI-R-CAO-bhe-c-02-0-UI"  
/clone\_lib="UI-R-CAO"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pRT73D-Pac (Pharmacia) with a modified  
polylinker. Site\_1: Not I; Site\_2: Eco RI; The UI-R-CAO  
library is a subtracted library derived from the following  
tissues: thalamus, cerebellum, hypothalamus, medulla, pons  
, midbrain, cerebral cortex, corpus striatum, testis, and  
hippocampus. For a detailed description of the library  
from which this clone was derived, please visit our web  
site at [ratest.eng.uiowa.edu](http://ratest.eng.uiowa.edu). The subtraction has been

BASE COUNT 101 a 106 c 114 g 116 t  
 ORIGIN previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)  
 TAG\_SEQ=None found"

Query Match 18.6%; Score 62.6; DB 148; Length 437;  
 Best Local Similarity 80.5%; Pred. No. 5.2e-09;  
 Matches 99; Conservative 0; Mismatches 16; Indels 8; Gaps 2;

OY 2 GCTGCATCTTTTCTATGCTCTCCCTGCTGCGCTGATGAGATACAGACGAAAACG 61  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 168 GCTGCATCTTTTCTATGCTCTCCCTGCTGCGCTGATGAGATACAGACGAAAACG 109  
 OY 62 GACAGCTGCTTCATGATGCTGCGACCCCTGCGTGCATGATGAGACGACCATATGTTG 121  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 108 GACAGCTGCTTCATGATGCTGCGACCCCTGCGTGCATGATGAGACGACCATATGTTG 57

OY 122 NAT 124  
 56 ATT 54

RESULT 10  
 BF394149/c 443 bp mRNA EST 27-NOV-2000  
 LOCUS BF394149  
 DEFINITION UI-R-CAO-bha-c-12-0-UI-s1 UI-R-CAO Rattus norvegicus cDNA clone  
 ACCESSION BF394149  
 VERSION BF394149.1 GI:11379013  
 KEYWORDS EST.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE 1 (bases 1 to 443)  
 AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 COMMENT Contact: Soares, MB  
 Program for Rat Gene Discovery and Mapping  
 University of Iowa  
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: msoares@blue.weeg.uiowa.edu  
 oligo-dt track not found. Not a site shown in beginning of sequence  
 is likely internal to the message. cDNA library Preparation: M.B.  
 Soares Lab Clone distribution: clones will be available through  
 Research Genetics (www.resgen.com)  
 Seq primer: M13 Forward  
 POLYA-No.

FEATURES  
 source Location/Qualifiers

1..443  
 /organism="Rattus norvegicus"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="UI-R-CAO-bha-c-12-0-UI"  
 /clone\_1db="UI-R-CAO"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-CAO  
 library is a subtracted library derived from the following  
 tissues: thalamus, cerebellum, hypothalamus, medulla, pons  
 and hippocampus. For a detailed description of the library  
 from which this clone was derived, please visit our web  
 site at ratseq.eng.uiowa.edu. The subtraction has been  
 previously described in (Bonaldo, Lennon and Soares,

BASE COUNT 102 a 106 c 117 g 118 t  
 ORIGIN Genome Research 6:791-806, 1996)  
 TAG\_SEQ=None found"

Query Match 18.6%; Score 62.6; DB 148; Length 443;  
 Best Local Similarity 80.5%; Pred. No. 5.2e-09;  
 Matches 99; Conservative 0; Mismatches 16; Indels 8; Gaps 2;

OY 2 GCTGCATCTTTTCTATGCTCTCCCTGCTGCGCTGATGAGATACAGACGAAAACG 61  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 174 GCTGCATCTTTTCTATGCTCTCCCTGCTGCGCTGATGAGATACAGACGAAAACG 115  
 OY 62 GACAGCTGCTTCATGATGCTGCGACCCCTGCGTGCATGATGAGACGACCATATGTTG 121  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 114 GACAGCTGCTTCATGATGCTGCGACCCCTGCGTGCATGATGAGACGACCATATGTTG 63

OY 122 NAT 124  
 62 ATT 60

RESULT 11  
 BG293757 711 bp mRNA EST 21-FEB-2001  
 LOCUS BG293757  
 DEFINITION 602390647F1 NIH\_MGC\_94 Mus musculus cDNA clone IMAGE:4502603 5',  
 mRNA sequence.  
 ACCESSION BG293757  
 VERSION BG293757.1 GI:13053739  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 711)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: gcapus@mail.nih.gov  
 Tissue Procurement: The Cepko Laboratory  
 cDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LHAM10371 row: n column: 12  
 High quality sequence stop: 708.

FEATURES  
 source Location/Qualifiers

1..711  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4502603"  
 /clone\_1db="NIH\_MGC\_94"  
 /tissue\_type="retina"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: eye; Vector: pCMV-Sport6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
 Average insert size 3.3 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC Library."  
 BASE COUNT 180 a 196 c 150 g 185 t  
 ORIGIN

Query Match 16.7%; Score 56; DB 175; Length 711;  
 Best Local Similarity 84.9%; Pred. No. 7.1e-07;  
 Matches 62; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
 OY 2 GCTGCATCTTTTCTATGCTCTCCCTGCTGCGCTGATGAGATACAGACGAAAACG 61  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: msources@blue.weeg.uiowa.edu  
 The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult 8-day Embryo library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics. This clone is also available through the I.M.A.G.E. Consortium at LNL (info@image.lnl.gov). IMAGE ID-1774109 The following repetitive elements were found in this cDNA sequence: 235-268, >(CAC)n(Simple\_repeat  
 Seq primer: M13 forward  
 POLYA-No.

# FEATURES

## SOURCE

1. 330  
 /location="Qualifiers  
 /organism="Rattus norvegicus"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="UI-R-C0-1e-a-08-0-01"  
 /clone\_1lb="UI-R-C0"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site\_1: Not I; Site\_2: Eco RI; The UI-R-C0 library is a subtracted library derived from the UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The UI-R-E1 library consisted of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C0) was constructed as follows: PCR amplified cDNA inserts from a pool of UI-R-A1 and UI-R-E1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the pooled UI-R-A1 and UI-R-E1 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C0 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)."

BASE COUNT 93 a 73 c 62 g 102 t  
 ORIGIN

Query Match 10.6%; Score 35.6; DB 14; Length 330;  
 Best Local Similarity 54.4%; Pred. No. 1.6;  
 Matches 68; Conservative 0; Mismatches 57; Indels 0; Gaps 0;  
 Oy 18 TGCCTCCCTGCGCTGATGGAGTACAGACGCAAGAGGACGCTGNTCATG 77  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 259 TGCCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 200  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Oy 78 ATTCAGTCGACCCCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 137  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 199 GGAGACTGGAACCTTTCAGACGACATGAGGACGAGGACGATTCCTCCAG 140  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Oy 138 CCGTT 142  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 139 ACTTT 135  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15  
 AA97859/c 519 bp mRNA EST 04-JUL-1999  
 LOCUS AA97859/c 519 bp mRNA EST 04-JUL-1999  
 DEFINITION UI-R-C0-hu-c-11-0-UI s1 UI-R-C0 Rattus norvegicus cDNA clone  
 UI-R-C0-hu-c-11-0-UI 3', mRNA sequence.

ACCESSION AA97859  
 VERSION AA97859.1 GI:4285835  
 KEYWORDS EST.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 519)  
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 COMMENT On Jun 5, 1998 this sequence version replaced gi:3188510.  
 Contact: Soares, MB  
 Program for Rat Gene Discovery and Mapping  
 University of Iowa  
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: msources@blue.weeg.uiowa.edu

The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult 8-day Embryo library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics. This clone is also available through the I.M.A.G.E. Consortium at LNL (info@image.lnl.gov). IMAGE ID-1773821 The following repetitive elements were found in this cDNA sequence: 235-268, >(CAC)n(Simple\_repeat  
 Seq primer: M13 forward  
 POLYA-No.

# FEATURES

## SOURCE

1. 519  
 /location="Qualifiers  
 /organism="Rattus norvegicus"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="UI-R-C0-hu-c-11-0-01"  
 /clone\_1lb="UI-R-C0"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site\_1: Not I; Site\_2: Eco RI; The UI-R-C0 library is a subtracted library derived from the UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The UI-R-E1 library consisted of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C0) was constructed as follows: PCR amplified cDNA inserts from a pool of UI-R-A1 and UI-R-E1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the pooled UI-R-A1 and UI-R-E1 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C0 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)."

BASE COUNT 146 a 118 c 119 g 136 t  
 ORIGIN

Query Match 10.6%; Score 35.6; DB 14; Length 519;  
 Best Local Similarity 54.4%; Pred. No. 1.8;  
 Matches 68; Conservative 0; Mismatches 57; Indels 0; Gaps 0;



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2001, 03:51:30 ; Search time 1192.14 seconds  
(without alignments)  
4359.525 Million cell updates/sec

Title: US-08-224-621-74

Perfect score: 336  
Sequence: 1 CGCTGCATCTTTTCTATGC.....CCCCNTTGTTCCACCA 336

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 773874588 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_ba3:\*  
4: gb\_in1:\*  
5: gb\_in2:\*  
6: gb\_in3:\*  
7: gb\_om:\*  
8: gb\_ov:\*  
9: gb\_pat1:\*  
10: gb\_pat2:\*  
11: gb\_ph:\*  
12: gb\_pi1:\*  
13: gb\_pi2:\*  
14: gb\_pi3:\*  
15: gb\_pi4:\*  
16: gb\_ba1:\*  
17: gb\_ba2:\*  
18: gb\_fun:\*  
19: em\_hcgo\_hum:\*  
20: em\_hcgo\_inv:\*  
21: em\_hcgo\_rnd:\*  
22: em\_hcgo\_hum1:\*  
23: em\_hcgo\_hum2:\*  
24: em\_hcgo\_hum3:\*  
25: em\_hcgo\_hum4:\*  
26: em\_hcgo\_hum5:\*  
27: em\_hcgo\_hum6:\*  
28: em\_hcgo\_hum7:\*  
29: em\_hcgo\_hum8:\*  
30: em\_hcgo\_inv1:\*  
31: em\_hcgo\_inv2:\*  
32: em\_hcgo\_other:\*  
33: em\_hcgo\_rnd:\*  
34: em\_hum1:\*  
35: em\_hum2:\*  
36: em\_hum3:\*  
37: em\_hum4:\*  
38: em\_hum5:\*  
39: em\_hum6:\*  
40: em\_hum7:\*  
41: em\_in:\*  
42: em\_om:\*  
43: em\_or:\*

44: em\_ov:\*  
45: em\_pat:\*  
46: em\_ph:\*  
47: em\_pi1:\*  
48: em\_pi2:\*  
49: em\_pi3:\*  
50: em\_pi4:\*  
51: em\_pi5:\*  
52: em\_pi6:\*  
53: em\_pi7:\*  
54: em\_pi8:\*  
55: em\_pi9:\*  
56: em\_pi10:\*  
57: em\_pi11:\*  
58: em\_pi12:\*  
59: em\_pi13:\*  
60: em\_pi14:\*  
61: em\_pi15:\*  
62: em\_pi16:\*  
63: em\_pi17:\*  
64: em\_pi18:\*  
65: em\_pi19:\*  
66: em\_pi20:\*  
67: em\_pi21:\*  
68: em\_pi22:\*  
69: em\_pi23:\*  
70: em\_pi24:\*  
71: em\_pi25:\*  
72: em\_pi26:\*  
73: em\_pi27:\*  
74: em\_pi28:\*  
75: em\_pi29:\*  
76: em\_pi30:\*  
77: em\_pi31:\*  
78: em\_pi32:\*  
79: em\_pi33:\*  
80: em\_pi34:\*  
81: em\_pi35:\*  
82: em\_pi36:\*  
83: em\_pi37:\*  
84: em\_pi38:\*  
85: em\_pi39:\*  
86: em\_pi40:\*  
87: em\_pi41:\*  
88: em\_pi42:\*  
89: em\_pi43:\*  
90: em\_pi44:\*  
91: em\_pi45:\*  
92: em\_pi46:\*  
93: em\_pi47:\*  
94: em\_pi48:\*  
95: em\_pi49:\*  
96: em\_pi50:\*  
97: em\_pi51:\*  
98: em\_pi52:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	23	6.8	1769	94	MMNDP	X83794 M.musculus
2	23	6.8	1774	94	MMNDP	X92397 M.musculus
3	22	6.5	1846	93	HSNDPG	X65882 H.sapiens m
4	22	6.5	1872	93	HSCHRX	X65724 H.sapiens D
5	22	6.5	119945	92	HSR218J18	AL034370 Human DNA
6	22	6.5	162409	67	AC022184	AC022184 Homo sapi
7	20	6.0	1493	91	BC006263	BC006263 Homo sapi
8	20	6.0	1792	89	AF236018	AF236018 Cercopit





/db\_xref="SWISS-PROT:P48744"  
/translation="MRNVLASISMLAIIMGDTSDSKTSSFLMDSORCMRHVYD  
SISHPLYKCSKMWLLARCEHCQSARSREPLVSFTVLKQPFRRSCHCCRPOTSKL  
ALRLRCSGGMRLATATRYTILSCHECCS"

exon  
/number=3  
600..1774

polya\_signal  
1752..1757

BASE COUNT 440 a 437 c 387 g 510 t  
ORIGIN

Query Match 6.8%; Score 23; DB 94; Length 1774;  
Best Local Similarity 100.0%; Pred. No. 0.029;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 38 ATGGAGATACAGACGCAAAAC 60  
|||||

Db 486 ATGGAGATACAGACGCAAAAC 508

RESULT 3  
LOCUS HSNDPG 1846 bp mRNA PRI 03-DEC-1993  
DEFINITION H.sapiens mRNA NDP.  
ACCESSION X65882.561557  
VERSION X65882.1 GI:35016  
KEYWORDS NDP gene.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 1846)  
Chen, Z.  
Direct Submission

JOURNAL Submitted (28-APR-1992) Z. Chen, University of Oxford, Genetics  
Lab/Dept of Biochemistry, South Parks Road, Oxford OX1 3QU, UK  
2 (bases 1 to 1846)

REFERENCE Chen, Z.Y., Hendriks, R.W., Jobling, M.A., Powell, J.F.,  
Breakfield, X.O., Sims, K.B. and Craig, I.W.  
Isolation and characterization of a candidate gene for Norrie  
disease

JOURNAL Nat. Genet. 1 (3), 204-208 (1992)  
MEDLINE 93265104

FEATURES  
source location/Qualifiers

1..1846

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="Xp11.23"

/dev\_stage="fetal"

/tissue\_type="retina"

/clone\_lib="fetal retina cDNA"

/clone="FR2, FR7"

85..92

409..810

/gene="NDP"

409..810

/gene="NDP"

/note="cDNA for Norrie Disease"

/codon\_start=1

/protein\_id="CAA46713.1"

/db\_xref="GI:35017"

/translation="MRNVLASISMLAIIMGDTSDSKTSSFLMDSORCMRHVY  
VDISHPLYKCSKMWLLARCEHCQSARSREPLVSFTVLKQPFRRSCHCCRPOTSK  
LALRLRCSGGMRLATATRYTILSCHECCS"

944..948

1811..1815

polya\_signal  
1811..1815

BASE COUNT 500 a 422 c 406 g 518 t  
ORIGIN

Query Match 6.5%; Score 22; DB 93; Length 1846;

Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 TTTCTATGCTCTCCCTGCTGG 32  
|||||

Db 436 TTTCTATGCTCTCCCTGCTGG 457

RESULT 4  
LOCUS HSCHRX 1872 bp mRNA PRI 05-AUG-1992  
DEFINITION H.sapiens DNA for ORF1 and ORF2 from chromosome X.  
ACCESSION X65724  
VERSION X65724.1 GI:29946  
KEYWORDS X chromosome.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 1872)  
Berger, W., Meindl, A., van de Pol, T.J., Cremers, F.P., Ropers, H.H.,  
Doerner, C., Monaco, A., Bergen, A.A., Lebo, R., Warburg, M. et al.  
Isolation of a candidate gene for Norrie disease by positional  
cloning  
Nat. Genet. 1 (3), 199-203 (1992)

JOURNAL 93265103  
MEDLINE Erratum: [[published erratum appears in Nat Genet 1992 Sep;2(1):84]]  
REMARK 2 (bases 1 to 1872)  
Berger, W.  
Direct Submission

JOURNAL Submitted (08-APR-1992) W. Berger, University Hospital Nijmegen,  
Dept of Human Genetics, Geert Grooteplein 20, P.O. box 9101, 6500 HB  
Nijmegen, THE NETHERLANDS

REFERENCE 1 (bases 1 to 1872)  
Berger, W.  
Direct Submission

JOURNAL Submitted (08-APR-1992) W. Berger, University Hospital Nijmegen,  
Dept of Human Genetics, Geert Grooteplein 20, P.O. box 9101, 6500 HB  
Nijmegen, THE NETHERLANDS

FEATURES  
source location/Qualifiers  
1..1872  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/clone\_lib="adult and fetal retina libr."  
/map="p11.4"  
417..818  
/note="ORF1"  
/codon\_start=1  
/protein\_id="CAA46639.1"  
/db\_xref="GI:29947"  
/translation="MRNVLASISMLAIIMGDTSDSKTSSFLMDSORCMRHVY  
VDISHPLYKCSKMWLLARCEHCQSARSREPLVSFTVLKQPFRRSCHCCRPOTSK  
LALRLRCSGGMRLATATRYTILSCHECCS"

727..1200  
/note="ORF2; Protein sequence is in conflict with the  
conceptual translation"  
/codon\_start=1  
/protein\_id="CAA46640.1"  
/db\_xref="GI:1335017"  
/translation="RHCGCDAGACDLSPTGTSSPYTARNAIPEARCCWLLDGTIV  
EAVRPARERLARLELROKMODFSRDSVSNDSYCLTERDTLTSTLQFSPSGT  
ISRGSRSDEGIFPLGSCCLGFGQFESISKNNALIKRVRLSVFGARMRITFGK"

CDS  
/note="ORF2; Protein sequence is in conflict with the  
conceptual translation"  
/codon\_start=1  
/protein\_id="CAA46640.1"  
/db\_xref="GI:1335017"  
/translation="RHCGCDAGACDLSPTGTSSPYTARNAIPEARCCWLLDGTIV  
EAVRPARERLARLELROKMODFSRDSVSNDSYCLTERDTLTSTLQFSPSGT  
ISRGSRSDEGIFPLGSCCLGFGQFESISKNNALIKRVRLSVFGARMRITFGK"

BASE COUNT 521 a 425 c 409 g 517 t  
ORIGIN

Query Match 6.5%; Score 22; DB 93; Length 1872;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 TTTCTATGCTCTCCCTGCTGG 32  
|||||

Db 444 TTTCTATGCTCTCCCTGCTGG 465

RESULT 5  
HSA218U18/c



```

/protein_id="CAA22268.1"
/db_xref="GI:3947697"
/translation="MRKHVLAASFMSLIVIGDTSKTDSFINDSPDRMRHHY
VDSISPLKSCSKVLLARCEHCSCASRSEPLVSEFVTKPFPSSCHCPCPORSK
LKLRLRCGSGMRLTFTYRIILSCHEECNS"
repeat_region
30396..30515
/note="MER5B repeat: matches 58..178 of consensus"
repeat_region
30652..30703
/note="26 copies 2 mer aa 73% conserved"
repeat_region
31354..31662
/note="AluB repeat: matches 1..308 of consensus"
repeat_region
32205..32377
/note="FAM repeat: matches 1..173 of consensus"
repeat_region
33007..33238
/note="L2 repeat: matches 2177..2417 of consensus"
repeat_region
33350..33381
/note="16 copies 2 mer ac 100% conserved"
misc_feature
complement(33360..33924)
/gene="NDP"
/feature="NDP"
repeat_region
33571..33727
/note="MER5A repeat: matches 9..189 of consensus"
repeat_region
33816..34022
/note="MIR repeat: matches 7..213 of consensus"
repeat_region
34667..34794
/note="L2 repeat: matches 2579..2698 of consensus"
repeat_region
35507..35714
/note="MIR repeat: matches 27..248 of consensus"
repeat_region
36253..36416
/note="MIR repeat: matches 63..244 of consensus"
repeat_region
40405..40533
/note="MIR repeat: matches 20..146 of consensus"
repeat_region
41308..41535
/note="L1P4 repeat: matches 5408..5640 of consensus"
prim_transcript
42540..43017
/feature="match: EST AA779268"
repeat_region
43414..43506
/feature="L2 repeat: matches 2646..2741 of consensus"
repeat_region
44406..44483
/feature="L1M4B repeat: matches 154..237 of consensus"
repeat_region
44932..45016
/feature="MIR repeat: matches 56..142 of consensus"
repeat_region
45355..45652
/feature="AluB repeat: matches 1..294 of consensus"
repeat_region
48165..48194
/feature="15 copies 2 mer aa 87% conserved"
misc_feature
48315..48514
/feature="match: STS AF020217"
repeat_region
51372..51487
/feature="L2 repeat: matches 2589..2709 of consensus"
repeat_region
53680..53741
/feature="31 copies 2 mer ag 90% conserved"
repeat_region
56812..57069
/feature="MIR repeat: matches 2..262 of consensus"

Query Match
Best Local Similarity 100.0%: Pred. No. 0.11:
Matches 22: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TTTTCTATGCTCTCCGCTGG 32
|||||
Db 38840 TTTTCTATGCTCTCCGCTGG 38819

RESULT 6
AC022184 162409 bp DNA HTG 10-SEP-2000
LOCUS Homo sapiens chromosome 4 clone RP11-5264 map 4, WORKING DRAFT
DEFINITION
SEQUENCE 12 unordered pieces.
AC022184
AC022184.3 GI:10047716
VERSION
AC022184.3
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
human.

```

```

ORGANISM
Homo sapiens
REFERENCE
AUTHORS
Mamakia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITILE
1 (bases 1 to 162409)
JOURNAL
Homo sapiens chromosome 4, clone RP11-5264
AUTHORS
Unpublished
2 (bases 1 to 162409)
REFERENCE
AUTHORS
Birren,B., Linton,L., Nussbaum,C. and Lander,E.
Anderson,S., Baldin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G., Castle,A.,
Choepeil,Y., Colangelo,M., Collins,S., Collins,M., Cooke,P.,
Dearlano,K., Dewar,K., Domino,M., Doyle,M., Fennestor,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczek,J., Levine,R., Liu,C., Liu,G., Locke,K.,
Macdonald,P., Margulis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheters,R., Meldrum,J., Meneses,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (26-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 10, 2000 this sequence version replaced gi:1707809.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2268
Center clone name: 52-G-4
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 150598 bases at least Q40
Consensus quality: 156229 bases at least Q30
Consensus quality: 158583 bases at least Q20
Insert size: 188000; agarose-fp
Insert size: 161309; sum-of-ctrls
Quality coverage: 3.7 in Q20 bases; agarose-fp
Quality coverage: 4.3 in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 6777: contig of 6777 bp in length
6778 6877: gap of 100 bp
6878 8136: contig of 1259 bp in length
8137 8236: gap of 100 bp
8237 8236: contig of 1760 bp in length
9997 10096: gap of 100 bp
10097 12552: contig of 2456 bp in length
12553 12652: gap of 100 bp
20027 20126: contig of 7374 bp in length
20127 20126: gap of 100 bp
20127 27970: contig of 7844 bp in length
27971 28070: gap of 100 bp
28071 36727: contig of 8657 bp in length

```

[illegible]

TITLE	Sirect Submission									
JOURNAL	Submitted (09-APR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA									
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a> Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: <a href="mailto:Robert_Strausberg@nih.gov">Robert_Strausberg@nih.gov</a> Tissue Procurement: DCTD/DTF cDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland: Web site: <a href="http://www.nisc.nih.gov/nisc_mgc@nihgri.nih.gov">http://www.nisc.nih.gov/nisc_mgc@nihgri.nih.gov</a> Contact: <a href="mailto:nisc_mgc@nihgri.nih.gov">nisc_mgc@nihgri.nih.gov</a> Shevchenko, Y., Wetherby, K.D., Bouffstrom, Sternberg, S.M., Benjamin, B., Blakesley, R.W., Beckett, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-T., Karlins, E., Legaspi, R., Lim, M., Maduro, O.L., Masello, C., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B., Stancu, P., Thompson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Zhang, L., H. and Green, E.D.									
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <a href="http://image.lnl.gov">http://image.lnl.gov</a> Series: IRAL Plate: 15 Row: 1 Column: 15. Location/Qualifiers 1. 1493 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:3949350" /tissue_type="Ovary, adenocarcinoma" /clone_id="NIH_MGC_9" /lab_host="DH10B-R" /note="vector: pOTB7" 1. 1212 /codon_start=1 /product="Similar to diacylglycerol O-acyltransferase (mouse) homolog" /protein_id="AAH06263.1" /db_xref="GI:13623319" /translation="GTRLMNCVYMLIISNARPLENIITKGIIVDPIQVSVLEPKDY SWPRLCIAANFVAFAFOVERKRLAAGALTEQGLLHVANATITICFPAAVLVE STIPGSLILMAHTITLFLKFSYRDVNSMCRARAKASAGKASAPHYVSYD NLATRYDLFEAPTLCELENIPEPRPIRKFLRLRIEMLEFLOLVGILQOMVPT IONMSRPFMDYSRILERILKLAIPNHLILFEYVLPFSCLNAVVELIQPGREYV RDNMSESVTFPMQNNIPVHKWCIHREYKPMILRRGSSKMMARPGVLAAPFREYIV SVPLRMRLMAFTGMAQIPLAMVGVGFGNGTGNAAVWLIIIGDPIAVLTVYDIT VLNTEAPAE"									
CDS										
BASE COUNT	269 a 481 c 407 g 336 t									
ORIGIN										
Query Match	6.0%: Score 20; DB 91; Length 1493;									
Best Local Similarity	100.0%; Pred. No. 1.7;									
Matches 20; Conservative	0; Mismatches 0; Indels 0; Gaps 0;									
QY	22	CTCCCTGCTGGCGCTGATGG	41							
Db	324	CTCCCTGCTGGCGCTGATGG	343							
RESULT	8									
AF236018										
LOCUS	AF236018 1792 bp mRNA PRI 12-AUG-2000									
DEFINITION	Cercopithecus aethiops diacyl-glycerol acyltransferase mRNA, complete cds.									
ACCESSION	AF236018									
VERSION	AF236018.1 GI:9796483									
KEYWORDS	African green monkey.									



REFERENCE 1 (bases 1 to 4237)  
AUTHORS Mertens, C., Kuhn, C. and Franke, W.W.  
TITLE Plakophilins 2a and 2b: constitutive proteins of dual location in the karyoplasm and the desmosomal plaque  
JOURNAL J. Cell Biol. 135 (4), 1009-1025 (1996)  
MEDLINE 97081101  
REFERENCE 2 (bases 1 to 4237)  
AUTHORS Mertens, C.  
TITLE Direct Submission  
JOURNAL Submitted (29-APR-1996) C. Mertens, German Cancer Research Center, Cellbiology, Im Neuenheimer Feld 280, Heidelberg, 69120, FRG  
REMARK 3 (bases 1 to 4237)  
AUTHORS Mertens, C.  
TITLE Direct Submission  
JOURNAL Submitted (09-JAN-1997) C. Mertens, German Cancer Research Center, Cellbiology, Im Neuenheimer Feld 280, Heidelberg, 69120, FRG  
COMMENT On Feb 8, 1997 this sequence version replaced gi:1322396.  
FEATURES  
SOURCE Location/Qualifiers  
1..4237  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
<1..>4237  
join(26..1402,1535..2671)  
/codon\_start=1  
/product="plakophilin 2a"  
/protein\_id="CA66265.1"  
/db\_xref="GI:1871541"  
/translation="MAAPGAPAEYGYIRTVLGGQITLGGDSSSLAPSEAKLKGSSGGGQTVKSLRTOEOVQOTLARKGRSSVNGNLHRTSSVPEVYVNLHVENFVGRSPVPTKTYDMLKAGTATYEGRWGRGTAQYSQKSVERSLRHLRLRLTSPSSPRAHVTHSDYQSORAGHTLHQSRRALVPRYARSELVGVSRAGTTSROHPTTHROYQHSQSVDFDSIPANPALLTPREPTSGMGNLEKEVYLTAGLTVGQVRLVLPLOPVONRASRSMSHOSFSTRTLRAGPSVAVDSQGRALTVGQAAQSGNLTERTFTDSQAGNADMEMLERAVSMLEADHMPRTISAATFIQHECFQSKARKRVNORGLTIKLIQKLVONEDYQRAVCGALRLVLEDNDKLEVLEAVPRLLOVLTOTRDIETKQITGLIAMSNSDKIKNMTTALLTLTNTIIPESGMEGEGYPRKNGVCI LHNLSYQLEAELEPEKSYONTYONRNITODNKSIGCFESRSRKYKEQIODEVPEEKSNPKGEVWLMHSIVIRMYLSLAKSVRNTQASIGALONLTAGSGMPETSVAGTVQKESGLQHKRMILVGDPSYKTAISLRNLSLONELAKETLPDLYSIIPTDVPSTDLLETTASACYT LNNIIONSYONARDLNTGCIQIMATISGDAVASKAKASVLLYSIMATHELHHA YKKAQFKKIDPFNSRTAKAYHSLKD"

CDS  
26..2671  
/codon\_start=1  
/product="plakophilin 2b"  
/protein\_id="CA66264.1"  
/db\_xref="GI:1871540"  
/translation="MAAPGAPAEYGYIRTVLGGQITLGGDSSSLAPSEAKLKGSSGGGQTVKSLRTOEOVQOTLARKGRSSVNGNLHRTSSVPEVYVNLHVENFVGRSPVPTKTYDMLKAGTATYEGRWGRGTAQYSQKSVERSLRHLRLRLTSPSSPRAHVTHSDYQSORAGHTLHQSRRALVPRYARSELVGVSRAGTTSROHPTTHROYQHSQSVDFDSIPANPALLTPREPTSGMGNLEKEVYLTAGLTVGQVRLVLPLOPVONRASRSMSHOSFSTRTLRAGPSVAVDSQGRALTVGQAAQSGNLTERTFTDSQAGNADMEMLERAVSMLEADHMPRTISAATFIQHECFQSKARKRVNORGLTIKLIQKLVONEDYQRAVCGALRLVLEDNDKLEVLEAVPRLLOVLTOTRDIETKQITGLIAMSNSDKIKNMTTALLTLTNTIIPESGMEGEGYPRKNGVCI LHNLSYQLEAELEPEKSYONTYONRNITODNKSIGCFESRSRKYKEQIODEVPEEKSNPKGEVWLMHSIVIRMYLSLAKSVRNTQASIGALONLTAGSGMPETSVAGTVQKESGLQHKRMILVGDPSYKTAISLRNLSLONELAKETLPDLYSIIPTDVPSTDLLETTASACYT LNNIIONSYONARDLNTGCIQIMATISGDAVASKAKASVLLYSIMATHELHHA YKKAQFKKIDPFNSRTAKAYHSLKD"

BASE COUNT 1234 a 1005 c 976 g 1022 t  
ORIGIN

Query Match 6.0%; Score 20; DB 93; Length 4237;  
Best Local Similarity 100.0%; Pred. No. 1.7;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 CTGCTGGCGCTGATGGAGA 45  
Db 1725 CTGCTGGCGCTGATGGAGA 1744  
RESULT 12  
LOCUS AF205589  
DEFINITION Homo sapiens chromosome 8 map 8q24.3-qter clone CTA-393612, WORKING DRAFT SEQUENCE, 12 unordered pieces.  
ACCESSION AF205589  
VERSION AF205589.2 GI:8151755  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 133075)  
AUTHORS Polley, A., Baumgart, C., Blechschmidt, K., Dette, M.D., Jahn, N., Menzel, U., Reichwald, K., Schilhabel, M.B., Schudy, A., Taudien, S., Wen, G., Siebert, R., Schlegelberger, B. and Rosenthal, A.  
TITLE Chromosome 8 genomic sequence  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 133075)  
AUTHORS Polley, A., Wen, G., Baumgart, C., Dette, M., Jahn, N., Schilhabel, M., Menzel, U. and Rosenthal, A.  
TITLE Direct Submission  
JOURNAL Submitted (27-OCT-1999) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstr. 11, Jena 07745, Germany  
AUTHORS 3 (bases 1 to 133075)  
TITLE Genome Sequencing Center Jena.  
COMMENT Direct Submission  
Submitted (24-MAY-2000) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstr. 11, Jena 07745, Germany  
On Jun 1, 2000 this working version replaced gi:6531668.  
NOTE: This is a 'working draft' sequence. It currently consists of 12 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as \* runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 16996: contig of 16996 bp in length  
\* 17096: gap of unknown length  
\* 17097  
\* 28105: contig of 11009 bp in length  
\* 28205: gap of unknown length  
\* 28206  
\* 44180: contig of 15975 bp in length  
\* 44181  
\* 44280: gap of unknown length  
\* 44281  
\* 68014: contig of 23734 bp in length  
\* 68114: gap of unknown length  
\* 68115  
\* 78199: contig of 10085 bp in length  
\* 78200  
\* 78299: gap of unknown length  
\* 78300  
\* 93263: contig of 14964 bp in length  
\* 93264  
\* 93364: gap of unknown length  
\* 93440: contig of 6077 bp in length  
\* 93441  
\* 99540: gap of unknown length  
\* 99541  
\* 104854: contig of 5314 bp in length  
\* 104855  
\* 104954: gap of unknown length  
\* 104955  
\* 108813: contig of 3859 bp in length  
\* 108814  
\* 108913: gap of unknown length  
\* 108914  
\* 112473: contig of 3560 bp in length  
\* 112474  
\* 112573: gap of unknown length  
\* 112574  
\* 114895: contig of 2322 bp in length  
\* 114896  
\* 114995: gap of unknown length  
\* 114996  
\* 133075: contig of 18080 bp in length.  
FEATURES  
source Location/Qualifiers  
1..133075  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="8 map 8q24.3-qter"  
/clone="CTA-393612"

misc.feature	1..16996	/note="assembly_fragment clone_end:SP6 vector_side:left" 11496..133075 /note="assembly_fragment clone_end:T7 vector_side:right"
BASE COUNT	26919 a 38778 c 39111 g 27166 t	1101 others
ORIGIN		
Query Match	6.0%:	Score 20: DB 78; Length 133075;
Best Local Similarity	100.0%:	Pred. NO. 1.6;
Matches	20: Conservative	0: Mismatches 0: Indels 0: Gaps 0:
OY	22 CTCCTGCTGGCGGTATG 41	
Dd	46580 CTCCTGCTGGCGGTATG 46599	
RESULT 13		
AC087588/c		
LOCUS		
DEFINITION	AC087588 137161 bp DNA HTG 18-FEB-2001	
ACCESSION	Homo sapiens chromosome 12 clone RP11-8P13B, WORKING DRAFT	
VERSION	SEQUENCE, 16 unordered pieces.	
KEYWORDS	AC087588	
SOURCE	AC087588_6 GI:12957592	
ORGANISM	HTG: HTGS_PHASE1: HTGS_DRAFT. human.	
REFERENCE	Homo sapiens	
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo. 1 (bases 1 to 137161)	
	Muzny D.M., Adams C., Adio-Ondola B., Ali-osman F.R., Allen C., Alsbrooks S.L., Amaratunga H.C., Are J.R., Banks T., Barria J., Benton J., Bimaga K., Blankenburg K., Bonnin D., Bouck J., Bowle S., Briteva M., Brown E., Brown M., Bryant N.P., Buhaay C., Burich P., Burkett C., Burrell K.L., Byrd N.C., Carrott T.F., Carter M., Cavazos S.R., Chacko J., Chavez D., Chan G., Chen R., Chen Z., Chowdhry I., Christopoulos C., Cleveland C.D., Cox C., Coyle M.D., Dathorne S.R., David R., Davila M.L., Davis C., Davy-Carroll L., Dederich D.A., Delaney K.R., Delgado O., Dem A.L., Ding Y., Dinh H.H., Doutheate K.J., Draper H., Dugan-Rocha S., Durbin K.J., Barnhart C., Edgar D., Edwards C.C., Elhaj C., Escotto M., Falls T., Ferraguto D., Flagg N., Ford J., Foster P., Franzen P., Gabisi A., Gao J., Garcia A., Garner T., Garza N., Gill R., Gorrell J.H., Guevara W., Gunaratne P., Hale S., Hamilton K., Harris C., Harris K., Hart M., Haylak P., Hawes A., Hernandez J., Hernandez O., Hodgson A., Hogues M., Hollaway C., Hollins B., Homs F., Howard S., Huber J., Huliy S., Hume J., Joudah S., Jacobson B., Jia Y., Johnson R., Jolivet S., Joudeh S., Karlsson E., Kelly S., Khan U., King L., Korvah J., Kovar C., Kravovic J., Kursh A., Landry N., Leaty B., Lewis L.C., Lewis L., Li J., Li Z., Lichtarge O., Lieu C., Liu J., Liu W., Loulsegh H., Lozano R.J., Lu X., Lucier A., Lucier R., Luna R., Ma J., Maheshwari M., Mapua P., Martin R., Martindale A., Martinez E., Massey E., McWhiney E., McLeod M.P., Meador M., Mel G., Metzger M., Miner G., Miner Z., Mitchell T., Mohabbat K., Morjan A., Morris S., Moser M., Neal D., Newton J., Newton N., Nguyen A., Nguyen N., Nguyen N., Nickerson E., Nwokwuwo S., Oguchi M., Okunolu G., Orangunye N., Oviedo R., Pace A., Payton B., Peery J., Perez L., Peters L., Pickens R., Plimus E., Pu L.L., Ruitz M., Ren Y., Rives M., Rojas A., Rojokkan I., Rolle M., Ruiz S., Saveley G., Scherer S., Scott T., Sparks A., Stanley H., Slison I., Sodergren E., Sonalke T., Spinks A., Stanely H., Stone H., Sutton A., Swalek A., Taber P., Tamerisa A., Tamerisa K., Tang H., Taney J., Taylor C., Taylor T., Telford B., Thomas N., Thomas S., Usmani K., Vasquez L., Vera V., Villalon D., Vinson R., Wall R., Wang S., Ward-Moore S., Warren R., Washington C., Watlington S., Williams G., Williamson A., Wleczyk R., Wooden S., Worley K., Wu C., Wu Y., Wu Y.F., Zhou J., Zortilla S., Nelson D. and Gibbs R.	

```

TITLE      Direct Submission
JOURNAL    Unpublished
REFERENCE   2 (bases 1 to 137161)
AUTHORS    Worley,K.C.
TITLE      Direct Submission
JOURNAL    Submitted (12-JAN-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            On Feb 17, 2001 this sequence version replaced gi:12831332.
COMMENT     ----- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ----- Project Information
            Center project name: HCVS
            Center clone name: RP11-8P13B
            ----- Summary Statistics
            Sequencing vector: ML3; 108821
            Chemistry: Dye-Primer Bodipy; 4% of reads
            Chemistry: Dye-Terminator Big Dye; 51% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 13116 bases at least Q40
            Consensus quality: 134949 bases at least Q30
            Consensus quality: 136756 bases at least Q20
            Estimated insert size: 133513; sum-of-contrigs estimation
            Quality coverage: 0x in Q20 bases; agarose-gel estimation
            Quality coverage: 5.1x in Q20 bases; sum-of-contrigs estimation
            -----
            * NOTE: Estimated insert size may differ from sequence length
            * (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html)
            * NOTE: This is a "working draft" sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contrigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            1
            23144: contig of 23144 bp in length
            *
            23145 23244: gap of unknown length
            *
            23245 51640: contig of 28396 bp in length
            *
            51641 51740: gap of unknown length
            *
            51741 66311: contig of 14571 bp in length
            *
            66312 66411: gap of unknown length
            *
            66412 82979: contig of 16568 bp in length
            *
            82980 83080 83079: gap of unknown length
            *
            83080 95663: contig of 12584 bp in length
            *
            95664 95763: gap of unknown length
            *
            95764 101213: contig of 5450 bp in length
            *
            101214 101313: gap of unknown length
            *
            101314 108548: contig of 7235 bp in length
            *
            108549 108648: gap of unknown length
            *
            108649 112966: contig of 4318 bp in length
            *
            112967 113066: gap of unknown length
            *
            113067 118426: contig of 5360 bp in length
            *
            118427 118526: gap of unknown length
            *
            118527 122670: contig of 4144 bp in length
            *
            122671 122770: gap of unknown length
            *
            122771 126345: contig of 3775 bp in length
            *
            126346 126645: gap of unknown length
            *
            126646 129310: contig of 2665 bp in length
            *
            129311 129410: gap of unknown length
            *
            129411 132396: contig of 2986 bp in length
            *
            132397 132496: gap of unknown length
            *
            132497 134604: contig of 2108 bp in length
            *
            134605 134704: gap of unknown length
            *
            134705 136040: contig of 1336 bp in length
            *
            136041 136140: gap of unknown length
            *
            136141 137161: contig of 1021 bp in length.
            -----
            Location/Qualifiers
            1..137161
            /organism="Homo sapiens"

```

/db.xref="taxon:9606"  
/chromosome="12"  
/clone="Rp11-8p13b"  
BASE COUNT 40772 a 27449 c 27320 g 40107 t 1513 others  
ORIGIN

Query Match 6.0%; Score 20; DB 77; Length 137161;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 26 CTGCTGGCCCTGATGGAGA 45  
|||||  
Db 101580 CTGCTGGCCCTGATGGAGA 101561

RESULT 14  
AC019269/c  
LOCUS AC019269/c  
DEFINITION Homo sapiens clone Rp11-8p13, WORKING DRAFT SEQUENCE, 10 unordered pieces.  
ACCESSION AC019269  
VERSION AC019269.4 GI:9966967  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 152511)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Homo sapiens, clone Rp11-8p13  
JOURNAL Unpublished  
RECORD 2 (bases 1 to 152511)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Beckert,R., Beda,F., Boguslavsky,L., Bouknight,B., Brown,A., Burkett,G., Castle,A., Chepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Deatellano,K., Dewar,K., Domino,M., Doyle,M., Feneator,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J., Gardyn,S., Grant,G., Hagos,B., Heatford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karacas,A., Klein,J., Landers,T., Lenoczky,J., Levine,R., Lien,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGuirk,A., McKernan,K., McPeeters,R., Meldrum,J., Menus,L., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K., Pierre,N., Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rotman,D., Roy,A., Santos,R., Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.  
DIRECT SUBMISSION  
Submitted (31-DEC-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 5, 2000 this sequence version replaced gi:7329377.  
All repeats were identified using RepeatMasker:  
Smt, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence.submissions@genome.wi.mit.edu  
Project Information  
Center project name: L2991  
Center clone name: 8\_P13  
Summary Statistics  
Sequencing vector: M13; M77815; 96% of reads  
Sequencing vector: Plasmid; n/a; 8-0.8% of reads  
4.40579710144928Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 148050 bases at least Q40  
Consensus quality: 150303 bases at least Q30

Consensus quality: 151107 bases at least Q20  
Insert size: 163000; agarose-ff  
Insert size: 151611; sum-of-contigs  
Quality coverage: 5.5 in Q20 bases; agarose-ff  
Quality cov.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 10 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1 8618: contig of 8618 bp in length  
\* 8619 8718: gap of 100 bp  
\* 8719 13037: contig of 4319 bp in length  
\* 13038 13137: gap of 100 bp  
\* 13138 19064: contig of 5927 bp in length  
\* 19065 19164: gap of 100 bp  
\* 19165 27544: contig of 8380 bp in length  
\* 27545 27644: gap of 100 bp  
\* 27645 36519: contig of 8875 bp in length  
\* 36520 36619: gap of 100 bp  
\* 36620 43866: contig of 7247 bp in length  
\* 43867 43966: gap of 100 bp  
\* 43967 59245: contig of 15279 bp in length  
\* 59246 59345: gap of 100 bp  
\* 59346 81568: contig of 22223 bp in length  
\* 81569 81668: gap of 100 bp  
\* 81669 110160: contig of 28492 bp in length  
\* 110161 110260: gap of 100 bp  
\* 110261 152511: contig of 42251 bp in length.  
Location/Qualifiers  
1..152511  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Rp11-8p13"  
/clone\_1lb="RPC1-11 Human Male BAC"  
1..8618  
/note="assembly-fragment  
clone\_end:sp6  
vector\_side:left"  
8719..13037  
/note="assembly-fragment"  
13138..19064  
/note="assembly-fragment"  
19165..27544  
/note="assembly-fragment"  
27645..36519  
/note="assembly-fragment"  
36620..43866  
/note="assembly-fragment"  
43967..59245  
/note="assembly-fragment"  
59346..81568  
/note="assembly-fragment"  
81669..110160  
/note="assembly-fragment"  
110261..152511  
/note="assembly-fragment"

BASE COUNT 45636 a 30476 c 30973 g 44525 t 901 others  
ORIGIN

Query Match 6.0%; Score 20; DB 65; Length 152511;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 26 CTGCTGGCCCTGATGGAGA 45  
|||||  
Db 131220 CTGCTGGCCCTGATGGAGA 131201





```

misc_feature      33764. 41183
                    /note="assembly_fragment"
misc_feature      41284. 49837
                    /note="assembly_fragment"
misc_feature      49938. 60570
                    /note="assembly_fragment"
misc_feature      60671. 72982
                    /note="assembly_fragment"
misc_feature      73083. 88506
                    /note="assembly_fragment"
misc_feature      88607. 103760
                    /note="assembly_fragment"
misc_feature      103861. 118920
                    /note="assembly_fragment"
misc_feature      119021. 135838
                    /note="assembly_fragment"
misc_feature      135939. 161694
                    /note="assembly_fragment"
BASE COUNT      49151 a 31116 c 31161 g 48366 t 1900 others
ORIGIN

```

```

Query Match      5.7%; Score 19; DB 71; Length 161694;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      10 TTTTCTATGCTCTCCCTG 28
        |||||
DB 33644 TTTTCTATGCTCTCCCTG 33626

```

Search completed: August 24, 2001, 04:34:56  
 Job time: 2606 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2001, 03:53:45 ; Search time 125.36 seconds  
(without alignments)  
1682.953 Million cell updates/sec

Title: US-08-224-621-74

Perfect score: 336

Sequence: 1 CGCTGCATCTTTTCTATGC.....CCCCNTTGTTCACCAACCA 336

Scoring table: OLIGO\_NUC

Gapop 60.0, Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

N.Geneseq\_0601:\*

- 1: /SIDSI/gcgdata/geneseq/geneseq/NA1980.DAT:\*
- 2: /SIDSI/gcgdata/geneseq/geneseq/NA1981.DAT:\*
- 3: /SIDSI/gcgdata/geneseq/geneseq/NA1982.DAT:\*
- 4: /SIDSI/gcgdata/geneseq/geneseq/NA1983.DAT:\*
- 5: /SIDSI/gcgdata/geneseq/geneseq/NA1984.DAT:\*
- 6: /SIDSI/gcgdata/geneseq/geneseq/NA1985.DAT:\*
- 7: /SIDSI/gcgdata/geneseq/geneseq/NA1986.DAT:\*
- 8: /SIDSI/gcgdata/geneseq/geneseq/NA1987.DAT:\*
- 9: /SIDSI/gcgdata/geneseq/geneseq/NA1988.DAT:\*
- 10: /SIDSI/gcgdata/geneseq/geneseq/NA1989.DAT:\*
- 11: /SIDSI/gcgdata/geneseq/geneseq/NA1990.DAT:\*
- 12: /SIDSI/gcgdata/geneseq/geneseq/NA1991.DAT:\*
- 13: /SIDSI/gcgdata/geneseq/geneseq/NA1992.DAT:\*
- 14: /SIDSI/gcgdata/geneseq/geneseq/NA1993.DAT:\*
- 15: /SIDSI/gcgdata/geneseq/geneseq/NA1994.DAT:\*
- 16: /SIDSI/gcgdata/geneseq/geneseq/NA1995.DAT:\*
- 17: /SIDSI/gcgdata/geneseq/geneseq/NA1996.DAT:\*
- 18: /SIDSI/gcgdata/geneseq/geneseq/NA1997.DAT:\*
- 19: /SIDSI/gcgdata/geneseq/geneseq/NA1998.DAT:\*
- 20: /SIDSI/gcgdata/geneseq/geneseq/NA1999.DAT:\*
- 21: /SIDSI/gcgdata/geneseq/geneseq/NA2000.DAT:\*
- 22: /SIDSI/gcgdata/geneseq/geneseq/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	6.0	1521	19	AAV01533
2	20	6.0	1895	21	AAAB8842
3	20	6.0	1895	21	AA45383
4	20	6.0	1976	21	AAV6169
5	16	4.8	144	19	AAV54158
6	16	4.8	150	22	AAV51166
7	16	4.8	300	21	AAA49869
8	16	4.8	507	20	AAV89118
9	16	4.8	528	20	AAV91501
10	16	4.8	559	20	AAV91502
11	16	4.8	1234	12	AA013722

12	16	4.8	1234	12	AA013723
13	16	4.8	1235	17	AAV40795
14	16	4.8	1235	17	AAV40796
15	16	4.8	1235	18	AAV91547
16	16	4.8	1235	18	AAV91548
17	16	4.8	2536	21	AAA47422
18	16	4.8	2773	20	AA222704
19	16	4.8	3056	21	AAV72111
20	16	4.8	5737	17	AAV12235
21	16	4.8	8894	21	AAV3169
22	16	4.8	10095	19	AAV58938
23	16	4.8	11122	22	AAV25342
24	15	4.5	21	22	AAV6549
25	15	4.5	274	21	AAV05618
26	15	4.5	388	21	AAV67571
27	15	4.5	447	21	AAV67559
28	15	4.5	459	21	AAV08947
29	15	4.5	477	21	AAV51731
30	15	4.5	488	21	AAV36861
31	15	4.5	488	21	AAV16466
32	15	4.5	495	21	AAV21226
33	15	4.5	495	21	AAV35104
34	15	4.5	513	16	AAV26145
35	15	4.5	673	21	AAV15181
36	15	4.5	711	21	AAV65990
37	15	4.5	858	21	AAV59981
38	15	4.5	978	21	AAV41725
39	15	4.5	1102	21	AAV66223
40	15	4.5	1117	21	AAV50473
41	15	4.5	1120	19	AAV05149
42	15	4.5	1122	21	AAV34669
43	15	4.5	1126	21	AAV297097
44	15	4.5	1128	21	AAV49032
45	15	4.5	1128	21	AAV73738

#### ALIGNMENTS

RESULT	1
ID	AAV01533 standard; DNA; 1521 BP.
AC	AAV01533;
XX	
DT	08-JUN-1998 (first entry)
XX	
DE	Human acylcoenzyme A:cholesterol acyltransferase II DNA.
XX	
KW	Acylcoenzyme A:cholesterol acyltransferase II; ARGP-1;
KW	ACAT related gene product 1; human; sterol esterification;
KW	inhibitor; atherosclerosis; hyperlipidaemia; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	Location/Qualifiers
FT	1..1224
FT	/tag= a
XX	
XX	WO9745439-A1.
PN	
PD	04-DEC-1997.
XX	
PF	30-MAY-1997; 97WO-US09460.
XX	
PR	30-MAY-1996; 96US-0657620.
XX	
PA	(UYCO ) UNIV COLUMBIA NEW YORK.
XX	
PI	Sturley SL;
XX	
DR	WPI; 1998-032573/03.
DR	P-PDB; AAV43406.

Lycopene cyclase -  
Lycopene cyclase c  
Recombinant lycopene  
Erwinia herbicola  
Genetically engineered  
Sequence encoding  
Pseudomonas fluorescens  
Human OREF ORF2766  
p60Pik cDNA. Mus  
Erwinia carotovora  
Mycobacterium tube  
Nucleotide sequenc  
Human gene single  
Human secreted pro  
Eucalyptus grandis  
Fusarium venenatum  
Zea mays DNA fragm  
Arabidopsis thaliana  
Human colon cancer  
Human low adenosin  
Human adenosin re  
Human gene signatu  
Trichoderma reesei  
E. coli proliferat  
Human secreted pro  
Arabidopsis thaliana  
cDNA encoding a ma  
Arabidopsis thaliana  
Nucleotide sequenc  
Arabidopsis thaliana  
Human secreted pro  
Human cardiac acti  
Human cardiac acti

XX DNA encoding acylcoenzyme A: cholesterol acyltransferase II or  
PT III - useful to identify inhibitors for treatment of  
PT atherosclerosis or hyperlipidaemia  
XX  
XX Claim 5; Fig 15A-B; 121pp; English.  
XX  
CC This isolated nucleic acid molecule encodes human acylcoenzyme  
CC A:cholesterol acyltransferase II (see AA043406), or ACAT related gene  
CC product 1 (ARGP-1), that is expressed at high levels in intestine  
CC and is a candidate for sterol esterification in these tissues. It  
CC was identified following database searching for human ACAT-related  
CC sequences and use of PCR and RACE to obtain full-length sequences.  
CC An isolated nucleic acid (see AA01534) for human ARGP-2 (see AA043407)  
CC was also identified. Also claimed are host vector systems for  
CC production of ARGP polypeptides, and transgenic non-human mammals.  
CC ARGP-1 and ARGP-2 nucleic acids can be used to diagnose or treat a  
CC subject who has an imbalance in sterol levels due to a defect in  
CC sterol esterification. A claimed oligonucleotide capable of  
CC specifically hybridising to a unique sequence of nucleotides in the  
CC isolated nucleic acid molecule, or a vector expressing the  
CC oligonucleotide, can be used to inhibit wild-type ARGP-1 or ARGP-2.  
CC The wild-type enzymes can be used to identify an inhibitor, useful  
CC in the treatment of atherosclerosis or hyperlipidaemia.  
XX  
SQ Sequence 1521 BP; 277 A; 489 C; 412 G; 343 T; 0 other;

Query Match 6.0%; Score 20; DB 19; Length 1521;  
Best Local Similarity 100.0%; Pred. No. 0.12;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 22 CTCCCTGCTGCGCGTGTATGG 41  
|||||  
DB 336 ctccctgctgctgctgctg 355

RESULT 2  
AAA88842  
ID AAA88842 standard; DNA; 1895 BP.  
XX  
AC AAA88842;  
XX  
DT 19-FEB-2001 (first entry)  
XX  
DE Human acyl CoA:cholesterol acyltransferase DNA.  
XX  
KW Acyl CoA:cholesterol acyltransferase; ACAT; transgenic plant;  
KW sterol; tocopherol; phytosterol; phytostanol; anticholesterolemic;  
KW hypolipemic; human; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200061771-A2.  
XX  
PD 19-OCT-2000.  
XX  
PF 12-APR-2000; 2000WO-US09696.  
XX  
PR 12-APR-1999; 99US-0128995.  
XX  
PA (MONS ) MONSANTO CO.  
XX  
PI Venkatramesh M, Corbin DR, Bhat GB, Boddupalli SS, Grebenok RJ;  
PI Kishore GM, Lardizabal KD, Lassner MW, Rangwala SH, Karunanandaa B;  
XX  
DR WPI: 2000-665136/64.  
XX  
PT Genetically engineering the biosynthetic pathways in plants involved in  
PT the accumulation of sterol compounds and tocopherol to produce  
PT compounds for lowering the level of low density lipoprotein cholesterol  
PT in blood serum -

PS Disclosure: Page 60-62; 166pp; English.  
XX  
XX The present sequence is that of human acyl CoA:cholesterol  
CC acyltransferase (ACAT) full-length DNA. Sterol O-acyltransferases  
CC such as ACAT catalyse the formation of cholesterol esters from  
CC cholesterol and long chain fatty acids. Recombinant constructs of  
CC the invention are used to alter the biosynthesis and accumulation  
CC of sterols and tocopherols in transgenic plants. Seeds of such  
CC plants may contain elevated levels of sitostanol and/or its esters,  
CC and alpha-tocopherol, and reduced levels of campesterol and  
CC campestanol and their esters. The seeds may also contain the novel  
CC sterol brassicstanol. Oil obtained from the seeds can be used in  
CC food and pharmaceutical compositions to lower levels of low density  
CC lipoprotein cholesterol in blood serum. ACAT enzymes can be used  
CC in the present invention to produce elevated levels of phytosterol  
CC and/or phytostanol esters.  
XX  
SQ Sequence 1895 BP; 323 A; 610 C; 555 G; 406 T; 1 other;

Query Match 6.0%; Score 20; DB 21; Length 1895;  
Best Local Similarity 100.0%; Pred. No. 0.12;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 22 CTCCCTGCTGCGCGTGTATGG 41  
|||||  
DB 524 ctccctgctgctgctgctg 543

RESULT 3  
AAZ45383  
ID AAZ45383 standard; DNA; 1895 BP.  
XX  
AC AAZ45383;  
XX  
DT 27-MAR-2000 (first entry)  
XX  
DE DNA encoding a protein related to ACAT-like proteins.  
XX  
KW Acyl-CoA:cholesterol acyltransferase; ACAT; ACAT-like protein; sterol;  
KW ester; triacylglycerol; fatty acyl-CoA; lipid composition; plant cell;  
KW diacylglycerol acyltransferase; DAGAT; triglyceride; cancer; diabetes;  
KW cardiopulmonary disease; heart failure; atherosclerosis; adipocytosis;  
KW leukaemia; skin carcinoma; fibroblastoma; metabolic disorder; obesity;  
KW abnormal lipid metabolism; abnormal fat absorption;  
KW lipoprotein secretion; adipogenesis; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO9963096-A2.  
XX  
PD 09-DEC-1999.  
XX  
PF 04-JUN-1999; 99WO-US12541.  
XX  
PR 05-JUN-1998; 98US-0088143.  
XX  
PR 12-NOV-1998; 98US-0108389.  
XX  
PA (CALJ ) CALGENE LLC.  
XX  
PI Lassner MW, Ruzlinsky DM;  
XX  
DR WPI: 2000-105701/09.  
XX  
PT Novel polynucleotides used for modifying plant oil composition and for  
PT developing products for treating e.g. cancer, diabetes, cardiopulmonary  
PT disease or metabolic disorders  
XX  
PS Claim 10; Fig 7; 89pp; English.  
XX  
CC The present sequence encodes a protein related to acyl-CoA:cholesterol  
CC acyltransferase (ACAT) related proteins. ACAT-like proteins are active  
CC in the formation of a sterol, ester and/or triacylglycerol from a fatty



Best Local Similarity 100.0%; Pred. No. 21;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 AGATACAGACAGCAA 58  
Db 58 agatacagacagcaa 73

## RESULT 6

AAAF95166  
ID AAF95166 standard; DNA; 150 BP.

AC AAF95166;

DT 23-MAY-2001 (first entry)

DE embB gene probe.

KW Tubercle bacillus; drug sensitivity; drug resistance; rifampicin;  
KW streptomycin; kanamycin; isoniazid; ethambutol; rpoB gene; rrs gene;  
XX rpsL gene; inhA gene; katG gene; embB gene; probe; PCR primer; ss.  
OS Mycobacterium tuberculosis.

PN EPI076099-A2.

PD 14-FEB-2001.

PF 02-AUG-2000; 2000EP-0306563.

PR 03-AUG-1999; 99JP-0220357.

PA (NISN) NISSHIMO IND INC.  
(SYST-) SYSTEM RES INC.

PI Suzuki Y, Nishida M, Takenishi S;

DR WPI; 2001-246696/26.

PT New oligonucleotides, nucleic acid probes and primers are useful for  
PT differentiating drug-resistance and determining infection with tubercle  
XX bacilli.

PS Disclosure; Page 64; 114pp; English.

CC The present invention relates to oligonucleotides based on nucleotide  
CC sequences obtained from both wild-type tubercle bacilli (WTB) that are  
CC susceptible to a drug and mutant-type tubercle bacilli (MTB) that are  
CC resistant to a drug. The drugs used in the present invention are  
CC rifampicin (RFP), streptomycin (SM), kanamycin (KM), isoniazid (INH) and  
CC ethambutol (EB). The rpoB gene is responsible for resistance to RFP; the  
CC rrs gene is responsible for resistance to SM and KM; the rpsL gene is  
CC responsible for resistance to SM; the inhA gene is responsible for  
CC resistance to INH; the katG gene is responsible for resistance to RFP;  
CC and the embB gene is responsible for resistance to EB. The present  
CC invention also relates to nucleic acid probes having part of a nucleotide  
CC sequence of tubercle bacilli (TB) responsible for drug resistance and  
CC primers used to generate the probes. The present sequence is an  
CC oligonucleotide of the present invention. The oligonucleotides of the  
CC present invention can be used to enable the differentiation of drug  
CC resistance and the determination of infection with tubercle bacilli  
CC simultaneously.

CC Sequence 150 BP; 26 A; 50 C; 43 G; 31 T; 0 other;

## Query Match

Best Local Similarity 4.8%; Score 16; DB 22; Length 150;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 CCTGCTGGCGCTGATG 40  
Db 125 cctgctggcgctgatg 140

## RESULT 7

AAA49869  
ID AAA49869 standard; DNA; 300 BP.

AC AAA49869;

DT 25-SEP-2000 (first entry)

DE Mycobacterium tuberculosis embB gene (ethambutol resistance).

KW Antibiotic resistance; embB gene; ethambutol resistance; ss.

OS Mycobacterium tuberculosis.

PH Key Location/Qualifiers

FT primer\_bind Complement(21..41)

FT primer\_bind /tag= "primer of AAA49847"

FT primer\_bind /note= "primer of AAA49847"

FT primer\_bind /tag= b

PN WO200036142-A1.

PD 22-JUN-2000.

PF 10-DEC-1999; 99WO-CA01177.

PR 11-DEC-1998; 98US-0111794.

PA (VISI-) VISIBLE GENETICS INC.

PI Shipman R;

DR WPI; 2000-431611/37.

PT Method for the detection and characterization of Mycobacterium  
PT tuberculosis with antibiotic resistance in a sample.

PS Disclosure; Page 8; 43pp; English.

CC The present sequence is that of the Mycobacterium tuberculosis  
CC embB (ethambutol resistance) gene (bp7741-8040). Amplification  
CC and cycle sequencing primers (see AAA49823-62) are used for the  
CC detection and analysis of antibiotic resistance-associated mutations  
CC in defined regions of rpoB (rifampin), katG (isoniazid), oxyR-aphC  
CC (streptomycin), mabA (isoniazid), rpsL/s12 (streptomycin), 16S/rrs  
CC (chloramphenicol), embA (ethambutol), pncA (pyrazinamide), gyrA  
CC These primers can be used in a method for the detection and  
CC characterization of M. tuberculosis present in a sputum sample.  
CC The method involves performing a sequencing procedure, with or  
CC without prior amplification, to detect the presence of M.  
CC tuberculosis, and if present to evaluate the rpoB, katG, rpsL/s12  
CC and 23S genes for the presence of antibiotic-inducing mutations.  
CC If M. tuberculosis is detected, a second sequencing procedure is  
CC performed on the sample to evaluate additional genes for the  
CC presence of antibiotic resistance-inducing mutations. Genotypic  
CC tests are rapid, sensitive and accurate providing information as to  
CC antibiotic treatment options.

CC Sequence 300 BP; 44 A; 101 C; 92 G; 63 T; 0 other;

## Query Match

Best Local Similarity 4.8%; Score 16; DB 21; Length 300;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 CCTGCTGGCGCTGATG 40  
Db 217 cctgctggcgctgatg 232

XX	RESULT	8
XX	AAV89118	
ID	AAV89118 standard; cDNA: 507 BP.	
AC	AAV89118;	
DT	15-FEB-1999 (first entry)	
XX	EST clone BZ288.	
DE		
XX	Human; secreted protein; expressed sequence tag; EST; haematopoiesis;	
KW	tissue growth; actinin; inhibin; chemotaxis; chemokinesis; haemostatic;	
KW	receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;	
RV	gene therapy; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	M09845436-A2.	
PD	15-OCT-1998.	
PF	10-APR-1998; 98WO-US06955.	
PR	10-APR-1997; 97US-0838821.	
PA	(GEMV ) GENETICS INST INC.	
PI	Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;	
PT	Racie LA, Spaulding V, Treacy M;	
WP	1999-070077/06.	
DR	New polynucleotides encoding human secreted proteins - derived from	
PT	e.g. human blood, kidney, foetal lung, placenta, testes, brain,	
PT	ovary, pituitary, retina and colon CDNA libraries.	
XX		
PS	Claim 1: Page 116; 618pp; English.	
XX	The present sequence represents a human expressed sequence tag (EST).	
CC	The polynucleotide, which is a secreted EST, and the encoded protein	
CC	are predicted to have useful biological activities which would make	
CC	them suitable for treating, preventing or ameliorating medical	
CC	conditions in humans and animals, although no supporting data is	
CC	given. Suggested activities include nutritional activity, immune	
CC	stimulating or suppressing activity, haematopoiesis regulating	
CC	activity, tissue growth activity, activin/inhibin activity,	
CC	chemotactic/hemokinetic activity, haemostatic and thrombolytic	
CC	activity, receptor/ligand activity, anti-inflammatory activity,	
CC	cadherin/tumour invasion suppressor activity, tumour inhibition	
CC	activity. The polynucleotide may also be useful for gene therapy.	
SQ	Sequence 507 BP; 122 A; 105 C; 101 G; 179 T; 0 other;	
XX		
XX	Query Match	4.8%; Score 16; DB 20; Length 507;
XX	Best Local Similarity	100.0%; Pred. No. 22;
Matches	16; Conservative	0; Mismatches
OY	166 AGAATAGTGGAGACTC	181
DB		
	173 agaatagtggagactct	188
XX		
XX	RESULT	9
XX	AAV91501	
ID	AAV91501 standard; DNA: 528 BP.	
AC	AAV91501;	
DT	28-SEP-1999 (first entry)	

DE		Cloned bovine herpes virus 2 thymidine kinase partial gene sequence.
XX		
KM	Bovine herpes virus-2; BHV-2; cytokine; immune response; immunogen; IL;	
KW	Interleukin; colony stimulating factor; CSF; Interferon; IFN; TNF;	
KW	tumour necrosis factor; recombinant virus; vaccine; mucosal disease;	
XX	mastitis; breast cancer; stress-induced disease; thymidine kinase; ss.	
OS	Bovine herpesvirus.	
PN	WO9916892-A1.	
PD	08-APR-1999.	
XX		
PF	29-SEP-1998; 98WO-GB029927.	
PR	29-SEP-1997; 97GB-0020633.	
XX	(UyBR-) UNIV BRISTOL.	
PA		
P1	Bradley AJ, Duffas WPH;	
XX		
DR	WPI; 1999-255101/21.	
DR	P-PSDB; AAY25085.	
XX		
P1	New bovine herpes virus-2 vectors	
XX		
PS	Example 2; Fig 4A-C; 130pp; English.	
CC	The invention provides bovine herpes virus-2 (BHV-2) based vectors that	
CC	comprise at least one cytokine-encoding DNA sequence. The expression of	
CC	Cytokines in mammals can up-regulate immune responses to the Immunogens.	
CC	The cytokine is selected from interleukins (IL), colony stimulating	
CC	factors (CSF), interferons (IFN) and tumour necrosis factors (TNF). The	
CC	BHV-2 based vector or recombinant virus can be used as vaccines. They can	
CC	be used for preventing or treating a mucosal disease in a subject, e.g.	
CC	ruminants, cattle, sheep, goats, dogs, cats, horses, or man, particularly	
CC	mastitis in cows or breast cancers in humans. They can also be used for	
CC	representing or treating a stress-induced disease. The present sequence	
CC	presents a BHV-2 thymidine kinase partial gene sequence, BHV-2UK cloned	
CC	for use in the invention. This differs from the previously published	
CC	sequence BHV-2NY-1 by having adenine at position 727 (full length)	
CC	instead of guanine.	
XX		
SQ	Sequence 528 BP; 114 A; 176 C; 151 G; 87 T; 0 other;	
	Query Match 4.8%; Score 16; DB 20; Length 528;	
	Best Local Similarity 100.0%; Pred. No. 22;	
	Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0.	
OY	24 CCCTGCTGGCGGCTGAT 39	
Db	241 ccctgctggcgtgat 256	
	RESULT 10	
ID	AAX91502	
XX	AAX91502 standard; DNA; 559 BP.	
AC	AAY91502;	
XX		
DT	28-SEP-1999 (first entry)	
XX		
DE	Bovine herpes virus 2 thymidine kinase partial gene sequence.	
XX		
KM	Bovine herpes virus-2; BHV-2; cytokine; immune response; immunogen; IL;	
KW	interleukin; colony stimulating factor; CSF; Interferon; IFN; TNF;	
KW	tumour necrosis factor; recombinant virus; vaccine; mucosal disease;	
XX	mastitis; breast cancer; stress-induced disease; thymidine kinase; ss.	
OS	Bovine herpesvirus.	
XX		
PN	WO9916892-A1	





PI Yen HC;  
 XX  
 DR MPI: 1991-281410/38.  
 DR P-PSDB: AAR13987.  
 XX  
 PT Biosynthesis of carotenoid(s) in genetically engineered hosts -  
 PT using DNA encoding enzymes from *Erwinia herbicola*  
 XX  
 PS Disclosure: Fig 19(1-3); 313pp: English.  
 XX  
 CC There are a total of six relevant genes in a 7900 bp region that  
 CC cause *E. coli* cells to produce GGPP and the carotenoids phytoene  
 CC through zeaxanthin digluconside, which is the final prod. identified  
 CC in the carotenoid pathway contd. In plasmid PARC376 (contg. a ca.  
 CC 13 kb chromosomal DNA fragment isolated by Perry et al., J. Bacteriol.,  
 CC 168:607 (1986). The genes for geranylgeranyl pyrophosphate (GGPP)  
 CC synthase, phytoene synthase, phytoene dehydrogenase-4H, lycopene  
 CC cyclase, beta-carotene hydroxylase, and zeaxanthin glycosylase are  
 CC represented in AAQ13716, AAQ13718, AAQ13719, AAQ13722, AAQ13724 and  
 CC AAQ13726 respectively.  
 CC The native sequence (AAQ13722) was genetically engineered for use in  
 CC expression in yeast. At the 5' end of the gene, the native  
 CC initiation GTG codon has been changed to an ATG codon. The second  
 CC amino acid residue, Arg, was originally encoded by an AGC codon  
 CC that was changed to a CGG codon, while retaining its coding for the  
 CC Arg amino acid residue.  
 CC Recombinant expression plasmids can be used to produce large amts.  
 CC of the enzymes and hence large amts. of the carotenoids which they  
 CC synthesize.  
 XX  
 SQ Sequence 1234 BP; 201 A; 386 C; 405 G; 242 T; 0 other;

Query Match 4.8%; Score 16; DB 12; Length 1234;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 125 TCTATCAGCTCACC CG 140  
 ||||||||||||  
 Db 896 tctatcagctcaccg 911

RESULT 13  
 AAT40795  
 ID AAT40795 standard; DNA: 1235 BP.  
 XX  
 AC AAT40795;  
 XX  
 DT 11-DEC-1996 (first entry)  
 XX  
 DE Lycopene cyclase coding sequence.  
 XX  
 KM GGPP synthase; lycopene cyclase; phytoene synthase; beta-carotene;  
 KM phytoene dehydrogenase-4H; carotenoid; C40; zeaxanthin; digluconside;  
 KM pigment; food colourant; chloroplast transit peptide; increase yield;  
 KM tobacco ribulose bis-phosphate carboxylase-oxygenase; ss.  
 XX  
 OS *Erwinia herbicola*.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 20..1144  
 FT /\*tag= a  
 FT /product= lycopene\_cyclase  
 XX  
 PN US5530188-A.  
 XX  
 PD 25-JUN-1996.  
 XX  
 PF 02-MAR-1990; 90US-0487613.  
 XX  
 PR 30-OCT-1991; 91US-0785566.  
 PR 02-MAR-1990; 90US-0487613.  
 PR 18-MAY-1990; 90US-0525351.

PR 03-AUG-1990; 90US-0562674.  
 PR 28-FEB-1991; 91US-0662921.  
 PR 21-JUL-1993; 91US-0095726.  
 XX  
 PA (STAD ) AMOCO CORP.  
 XX  
 PI Ausich RL, Brinkhaus FL, Mukharji I, Proffitt J;  
 PI Yarger J, Yen HB;  
 XX  
 DR MPI: 1996-308823/31.  
 DR P-PSDB: AAM01125.  
 XX  
 PT Increasing prodn. of total carotenoid(s) in a higher plant - by  
 PT transforming with vector encoding chloroplast transit peptide  
 PT operably linked to the *Erwinia herbicola* lycopene cyclase structural  
 PT gene  
 XX  
 PS Example 17; Column 99-102; 99pp: English.  
 XX  
 CC Manipulation of the present sequence, the lycopene cyclase coding  
 CC sequence, by in frame linkage to the chloroplast transit peptide  
 CC (AAM01124) of the tobacco ribulose bis-phosphate carboxylase-oxygenase  
 CC gene can lead to increased production of total carotenoids in the  
 CC chloroplast of transformed plants as compared to native, non-transformed  
 CC plants of the same type. Beta-carotene is an effective and apparently  
 CC harmless food colourant and is also in the pathway for biological  
 CC synthesis of further C40 carotenoids such as zeaxanthin and zeaxanthin  
 CC digluconside. Other enzymes involved in the carotenoid biosynthesis  
 CC pathway include geranylgeranyl pyrophosphate (AAM01119), phytoene  
 CC synthase (WO1121) and phytoene dehydrogenase-4H (AAM01123).  
 XX  
 SQ Sequence 1235 BP; 202 A; 384 C; 405 G; 244 T; 0 other;

Query Match 4.8%; Score 16; DB 17; Length 1235;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 125 TCTATCAGCTCACC CG 140  
 ||||||||||||  
 Db 897 tctatcagctcaccg 912

RESULT 14  
 AAT40796  
 ID AAT40796 standard; DNA: 1235 BP.  
 XX  
 AC AAT40796;  
 XX  
 DT 11-DEC-1996 (first entry)  
 XX  
 DE Recombinant lycopene cyclase coding sequence.  
 XX  
 KM GGPP synthase; lycopene cyclase; phytoene synthase; beta-carotene;  
 KM phytoene dehydrogenase-4H; carotenoid; C40; zeaxanthin; digluconside;  
 KM pigment; food colourant; chloroplast transit peptide; increase yield;  
 KM tobacco ribulose bis-phosphate carboxylase-oxygenase; ss.  
 XX  
 OS *Erwinia herbicola*.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 20..1144  
 FT /\*tag= a  
 FT /product= lycopene\_cyclase  
 FT misc\_feature 18..1165  
 FT /\*tag= b  
 FT /note= "SphI-BamHI fragment; claim 6"  
 XX  
 PN US5530188-A.  
 XX  
 PD 25-JUN-1996.  
 XX  
 PF 02-MAR-1990; 90US-0487613.

```

XX 30-OCT-1991; 91US-0785566.
PR 02-MAR-1990; 90US-0487613.
PR 18-MAY-1990; 90US-0525551.
PR 03-AUG-1990; 90US-0562674.
PR 28-FEB-1991; 91US-0662921.
PR 21-JUL-1993; 93US-0095726.
XX (STAD ) AMOCO CORP.
XX Ausich RL, Brinkhaus FL, Mukharji I, Proffitt J;
PI Yarger J, Yen HB;
XX WPI; 1996-308823/31.
DR P-PSDB; AAW01125.
XX
PT Increasing prodn. of total carotenoid(s) in a higher plant - by
PT transforming with vector encoding chloroplast transit peptide
PT operably linked to the Erwinia herbicola lycopene cyclase structural
PT gene
XX
PS Example 17; Fig 19; 99pp; English.
XX
CC The present sequence is a recombinant lycopene cyclase coding sequence.
CC The approx. 1142 bp SphI-BamHI fragment present in the plasmid PARC1509
CC (ATCC 40850) is used to transform plants in order to increase the level
CC of carotenoids produced by plants. Manipulation of it by in frame
CC linkage to the chloroplast transit peptide (AAW01124) of the tobacco
CC ribulose bis- phosphate carboxylase-oxygenase gene can lead to increased
CC production of total carotenoids in the chloroplast of transformed plants
CC as compared to native, non-transformed plants of the same type.
CC Beta-carotene is an effective and apparently harmless food colourant and
CC is also in the pathway for biological synthesis of further C40
CC carotenoids such as zeaxanthin and zeaxanthin diglycoside. Other enzymes
CC involved in the carotenoid biosynthesis pathway include geranylgeranyl
CC pyrophosphate (AAW01119), phytoene synthase (W01121) and phytoene
CC dehydrogenase-4H (AAW01123).
XX
SQ Sequence 1235 BP; 198 A; 388 C; 406 G; 243 T; 0 other;

Query Match          4.8%; Score 16; DB 17; Length 1235;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 TCTATCAGCTCACCG 140
   |||||||
DB 897 tctatcagctcacccg 912

RESULT 15
AAT91547
ID AAT91547 standard; DNA; 1235 BP.
XX
AC AAT91547;
XX
DT 15-JAN-1998 (first entry)
XX
DE Erwinia herbicola lycopene cyclase structural gene.
XX
KW Geranylgeranyl pyrophosphate synthase; Erwinia herbicola; GGP;
KW lycopene cyclase; recombinant enzyme; transgenic organism;
KW beta-carotene; yeast; plant; vitamin A; cancer; ss.
XX
OS Erwinia herbicola.
XX
FH Key Location/Qualifiers
FT CDS 20..1144
FT /*tag= a
FT /product= lycopene cyclase
FT /transl_except= (pos:20..22,aa:Met)
XX
XX US5656472-A.

```

```

XX 12-AUG-1997.
PD 02-MAR-1990; 90US-0487613.
XX
XX 21-JUL-1993; 93US-0095726.
XX 02-MAR-1990; 90US-0487613.
PR 18-MAY-1990; 90US-0525551.
PR 03-AUG-1990; 90US-0562674.
PR 28-FEB-1991; 91US-0662921.
PR 07-JUN-1995; 95US-0473512.
XX (STAD ) AMOCO CORP.
XX
XX Ausich RL, Brinkhaus FL, Mukharji I, Proffitt J;
PI Yarger J, Yen HB;
XX WPI; 1997-414592/38.
DR P-PSDB; AAW32474.
XX
PT DNA encoding Erwinia herbicola lycopene cyclase - for producing
PT recombinant enzyme; and transgenic organisms with increased
PT beta-carotene levels
XX
PS Example 17; Fig 19; 102pp; English.
XX
CC A novel DNA molecule has been isolated which encodes an Erwinia
CC herbicola lycopene cyclase enzyme that converts lycopene to beta-
CC carotene. The DNA molecule comprises at least 1125 bp and is present
CC in the plasmids PARC147, PARC1509, PARC1510 and PARC1520. The present
CC sequence represents the structural gene for lycopene cyclase. The
CC new DNA molecule can be used to produce the recombinant enzyme and
CC transgenic organisms, e.g. yeasts or plants, with increased beta-
CC carotene levels. Beta-carotene is used as a colourant in margarine and
CC butter and as an intermediate for vitamin A, and may prevent cancer.
XX
SQ Sequence 1235 BP; 202 A; 384 C; 405 G; 244 T; 0 other;

Query Match          4.8%; Score 16; DB 18; Length 1235;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 TCTATCAGCTCACCG 140
   |||||||
DB 897 tctatcagctcacccg 912

Search completed: August 24, 2001, 04:36:49
Job time: 2584 sec

```

This Page Blank (uspto)

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**  
**Best Available Copy**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2001, 03:51:39 ; Search time 57.47 Seconds

(Without alignments)  
1106.813 Million cell updates/sec

Title: US-08-224-621-74

Perfect score: 336  
Sequence: 1 CGCTGCATCTTTCTATGC.....CCCCNTGTTCCCAACCA 336

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 324599 seqs, 9465562 residues

Word size : 0

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/lna/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/lna/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/lna/5A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/lna/5B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/lna/PCRTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/lna/backfillseq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	6.0	1976	3	US-09-165-042-2
2	16	4.8	1235	1	US-08-095-726-13
3	16	4.8	1235	1	US-08-095-726-13
4	16	4.8	1235	1	US-08-095-726-13
5	16	4.8	1235	1	US-08-096-623A-15
6	16	4.8	5737	1	US-08-259-264-1
7	16	4.8	10095	3	US-08-822-586-45
8	15	4.5	1120	1	US-08-203-806B-3
9	15	4.5	1128	3	US-09-106-217-15
10	15	4.5	1134	3	US-09-106-217-1
11	15	4.5	1350	3	US-08-462-351-1
12	15	4.5	1350	6	5468481-2
13	15	4.5	1350	6	5194425-2
14	15	4.5	1488	2	US-08-812-203-4
15	15	4.5	1488	4	US-09-300-864-4
16	15	4.5	1728	3	US-08-379-802-1
17	15	4.5	1728	3	US-09-048-129-1
18	15	4.5	1728	4	US-09-048-079-1
19	15	4.5	1839	1	US-08-442-248-3
20	15	4.5	1839	1	US-08-440-815-3
21	15	4.5	1881	3	US-09-235-246-2
22	15	4.5	2022	1	US-08-803-973-6
23	15	4.5	2022	1	US-08-803-972-6
24	15	4.5	2065	3	US-08-335-8650-8
25	15	4.5	2124	1	US-08-803-973-11
26	15	4.5	2124	1	US-08-803-972-11
27	15	4.5	3336	3	US-09-330-970-4

28	15	4.5	3697	1	US-08-571-758-1	Sequence 1, Appl
29	15	4.5	3697	1	US-08-909-984A-1	Sequence 1, Appl
30	15	4.5	3697	1	US-08-909-983-1	Sequence 1, Appl
31	15	4.5	4371	1	US-08-803-973-1	Sequence 1, Appl
32	15	4.5	4371	1	US-08-803-972-1	Sequence 1, Appl
33	15	4.5	43795	3	US-08-742-185-101	Sequence 101, App
34	14	4.2	37	1	US-08-484-686B-60	Sequence 60, Appl
35	14	4.2	37	4	US-08-463-160B-60	Sequence 60, Appl
36	14	4.2	37	5	PCT-US91-02568-18	Sequence 18, Appl
37	14	4.2	297	4	US-09-060-756-574	Sequence 574, App
38	14	4.2	309	1	US-08-086-410-24	Sequence 24, Appl
39	14	4.2	330	5	PCT-US95-08596-1	Sequence 1, Appl
40	14	4.2	356	4	US-09-060-756-396	Sequence 396, App
41	14	4.2	359	3	US-08-589-028-3	Sequence 3, Appl
42	14	4.2	359	3	US-08-784-582-3	Sequence 3, Appl
43	14	4.2	359	4	US-08-785-271-3	Sequence 3, Appl
44	14	4.2	360	1	US-07-920-519-28	Sequence 28, Appl
45	14	4.2	360	1	US-08-086-410-21	Sequence 21, Appl

#### ALIGNMENTS

RESULT 1  
US-09-165-042-2  
; Sequence 2, Application US/09165042  
; Patent No. 6100077  
; GENERAL INFORMATION:  
; APPLICANT: Sturley, Stephen L.  
; APPLICANT: Oelkers, Peter  
; TITLE OF INVENTION: ISOLATION OF A GENE ENCODING DIACYLGLYCEROL  
; TITLE OF INVENTION: ACYLTRANSFERASE  
; FILE REFERENCE: 0575/56331  
; CURRENT APPLICATION NUMBER: US/09/165,042  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 1976  
; TYPE: DNA  
; ORGANISM: Yeast  
; US-09-165-042-2

Query Match 6.0%; Score 20; DB 3; Length 1976;

Best Local Similarity 100.0%; Pred. No. 0.018; Mismatches 0; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 22 CTCCTGCTGCGCTGATGG 41  
|||||  
DB 823 CTCCTGCTGCGCTGATGG 842

RESULT 2  
US-08-095-726-13  
; Sequence 13, Application US/08095726  
; Patent No. 5530188  
; GENERAL INFORMATION:  
; APPLICANT: Ausich, Rodney L  
; APPLICANT: Brinkhaus, Friedhelm L  
; APPLICANT: Mukharji, Indranil  
; APPLICANT: Proffitt, John H  
; APPLICANT: Yarger, James G  
; APPLICANT: Yen, Huel-Che B  
; TITLE OF INVENTION: Beta-Carotene Biosynthesis in  
; TITLE OF INVENTION: Genetically Engineered Hosts  
; NUMBER OF SEQUENCES: 79  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amoco Corp., Patents and Licensing Dept  
; STREET: 200 E Randolph St  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: USA

ZIP: 60680-0703  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/095,726  
FILING DATE: 21-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/785,566  
FILING DATE: 30-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Galloway, No. 5530188val B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 3128567180  
TELEFAX: 3128564972  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1235 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-095-726-13

Query Match 4.8%; Score 16; DB 1; Length 1235;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 TCTATCAGCTCACCG 140  
|||||  
DB 897 TCTATCAGCTCACCG 912

RESULT 3  
US-08-095-726-15  
Sequence 15, Application US/08095726  
Patent No. 5530188  
GENERAL INFORMATION:  
APPLICANT: Ausich, Rodney L.  
APPLICANT: Brinkhaus, Friedhelm L.  
APPLICANT: Mukharji, Indrani  
APPLICANT: Proffitt, John H.  
APPLICANT: Yarger, James G.  
APPLICANT: Yen, Huei-Che B.  
TITLE OF INVENTION: Beta-Carotene Biosynthesis in  
Genetically Engineered Hosts  
NUMBER OF SEQUENCES: 79  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amoco Corp., Patents and Licensing Dept  
STREET: 200 E Randolph St  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60680-0703  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/095,726  
FILING DATE: 21-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/785,566  
FILING DATE: 30-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Galloway, No. 5530188val B  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 3128567180  
TELEFAX: 3128564972  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1235 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-095-726-15

Query Match 4.8%; Score 16; DB 1; Length 1235;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 TCTATCAGCTCACCG 140  
|||||  
DB 897 TCTATCAGCTCACCG 912

RESULT 4  
US-08-096-623A-13  
Sequence 13, Application US/08096623A  
Patent No. 5684238  
GENERAL INFORMATION:  
APPLICANT: Ausich, Rodney L.  
APPLICANT: Brinkhaus, Friedhelm L.  
APPLICANT: Mukharji, Indrani  
APPLICANT: Proffitt, John H.  
APPLICANT: Yarger, James G.  
APPLICANT: Yen, Huei-Che B.  
TITLE OF INVENTION: Biosynthesis of zeaxanthin and  
glycosylated zeaxanthin in Genetically Engineered Hosts  
NUMBER OF SEQUENCES: 104  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Welsh & Katz, Ltd.  
STREET: 120 S. Riverside Plaza, 22nd Floor  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/096,623A  
FILING DATE: 22-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/805,061  
FILING DATE: 09-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/662,921  
FILING DATE: 28-FEB-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/562,674  
FILING DATE: 03-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/525,551  
FILING DATE: 18-MAY-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/487,613  
FILING DATE: 02-MAR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Gamson, Edward P.  
REGISTRATION NUMBER: 29,381  
REFERENCE/DOCKET NUMBER: AMO-006.1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 655-1500  
TELEFAX: (312) 655-1501

INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1235 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
POSITION IN GENOME:  
MAP POSITION: -19 to 1216  
UNITS: bp  
US-08-096-623A-13

Query Match 4.8%; Score 16; DB 1; Length 1235;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 TCTATCAGCTCACC CG 140  
DB 897 TCTATCAGCTCACC CG 912

## RESULT 5

US-08-096-623A-15  
Sequence 15, Application US/08096623A  
Patent No. 5684238  
GENERAL INFORMATION:  
APPLICANT: Ausich, Rodney L.  
APPLICANT: Brinkhaus, Friedhelm L.  
APPLICANT: Mukharji, Indranil  
APPLICANT: Proffitt, John H.  
APPLICANT: Yarger, James G.  
APPLICANT: Yen, Huel-Che B.  
TITLE OF INVENTION: Biosynthesis of Zeaxanthin and  
TITLE OF INVENTION: Glycosylated Zeaxanthin in Genetically Engineered Hosts  
NUMBER OF SEQUENCES: 104  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Welsh & Katz, Ltd.  
STREET: 120 S. Riverside Plaza, 22nd Floor  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/096,623A  
FILING DATE: 22-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/805,061  
FILING DATE: 09-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/662,921  
FILING DATE: 28-FEB-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/562,674  
FILING DATE: 03-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/525,551  
FILING DATE: 18-MAY-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/487,613  
FILING DATE: 02-MAR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Gansson, Edward P.  
REGISTRATION NUMBER: 29,381  
REFERENCE/DOCKET NUMBER: AMO-006.1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 655-1500

TELEFAX: (312) 655-1501  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1235 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
POSITION IN GENOME:  
MAP POSITION: -19 to 1216  
UNITS: bp  
US-08-096-623A-15

Query Match 4.8%; Score 16; DB 1; Length 1235;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 TCTATCAGCTCACC CG 140  
DB 897 TCTATCAGCTCACC CG 912

## RESULT 6

US-08-259-264-1/C  
Sequence 1, Application US/08259264  
Patent No. 5650293  
GENERAL INFORMATION:  
APPLICANT: White, Morris F.  
TITLE OF INVENTION: p66DPR: A DOWNSTREAM ELEMENT IN INSULIN SIGNALING  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 STATE STREET, SUITE 510  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/259,264  
FILING DATE: 10-JUN-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Myers, Louis  
REGISTRATION NUMBER: 35,965  
REFERENCE/DOCKET NUMBER: JDP-021  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5737 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1388..2749  
US-08-259-264-1

Query Match 4.8%; Score 16; DB 1; Length 5737;  
Best Local Similarity 100.0%; Pred. No. 4.6;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 17 ATGCTTCCTGCTGG 32  
|||||

DB 374 ATGCTCTCCCTGCTG 359

## RESULT 7

US-08-822-586-45  
; Sequence 45, Application US/08822586  
; Patent No. 6015890

## GENERAL INFORMATION:

APPLICANT: WILLIAM R. JACOBS, JR., JAMES M. MUSSER AND  
APPLICANT: AMALIO TELENITI  
TITLE OF INVENTION: AN EMBCAB OPERON OF MYCOBACTERIA AND  
TITLE OF INVENTION: MUTANTS THEREOF  
NUMBER OF SEQUENCES: 57  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN  
STREET: 90 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: U.S.A.  
ZIP: 10016

## COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 INCH 1.44 Mb STORAGE  
MEDIUM TYPE: DISKETTE  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/822,586  
FILING DATE: MARCH 20, 1997

## ATTORNEY/AGENT INFORMATION:

NAME: ELIZABETH A. BOGOSIAN  
REGISTRATION NUMBER: 39,911  
REFERENCE/DOCKET NUMBER: 96700/437  
TELEPHONE: (212) 697-5995

TELEFAX: (212) 286-0854 or 286-0082

TELEX: TWX 710-581-4766

INFORMATION FOR SEQ ID NO: 45:

## SEQUENCE CHARACTERISTICS:

LENGTH: 10095  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
HYPOTHETICAL: NO  
US-08-822-586-45

Query Match 4.8%; Score 16; DB 3; Length 10095;  
Best Local Similarity 100.0%; Pred. No. 4.8;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 CCTGCTGCGCTGTGANG 40

DB 7803 CCTGCTGCGCTGTGATG 7818

## RESULT 8

US-08-203-806B-3/C  
; Sequence 3, Application US/08203806B  
; Patent No. 5714575

## GENERAL INFORMATION:

APPLICANT: Inouye, Masayori  
APPLICANT: Jones, Pamela  
APPLICANT: Etchegaray, Jean-Pierre  
APPLICANT: Weining, Jlan  
APPLICANT: Pollitt, N. Stephen  
APPLICANT: Goldstein, Joel

TITLE OF INVENTION: Nucleic Acid Sequences, Stress-induced  
NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: Weiser & Associates

STREET: 230 South Fifteenth Street, Suite 500  
CITY: Philadelphia

STATE: PA  
COUNTRY: U.S.A.  
ZIP: 19102

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/203,806B  
FILING DATE: 01-MAR-1994  
CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Weiser, Gerard J.  
REGISTRATION NUMBER: 19,763  
REFERENCE/DOCKET NUMBER: 377,5998P

TELEPHONE: 215-875-8383

TELEFAX: 215-875-8394

TELEX: 834809 WEISSTAK

INFORMATION FOR SEQ ID NO: 3:

## SEQUENCE CHARACTERISTICS:

LENGTH: 1120 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 686..895  
US-08-203-806B-3

Query Match 4.5%; Score 15; DB 1; Length 1120;  
Best Local Similarity 100.0%; Pred. No. 16;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 CTGCATGAGGCACCA 113

DB 622 CTGCATGAGGCACCA 608

## RESULT 9

US-09-106-217-15  
; Sequence 15, Application US/09106217  
; Patent No. 6063576

## GENERAL INFORMATION:

APPLICANT: Keating, Mark T.  
APPLICANT: Olson, Timothy M.  
TITLE OF INVENTION: Actin Mutations in Dilated  
CARDIOMYOPATHY, a Heritable Form of Heart Failure  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.  
STREET: 555 Thirteenth Street, N.W., Suite 701 East  
CITY: Washington

STATE: DC

COUNTRY: U.S.A.

ZIP: 20004

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/106,217  
FILING DATE:

## CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Saxe, Stephen A.



REGISTRATION NUMBER: 38,609  
REFERENCE/DOCKET NUMBER: 2323-125  
TELEPHONE: 202-783-6040  
TELEFAX: 202-783-6031  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1128 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1125  
US-09-106-217-15

Query Match 4.5%; Score 15; DB 3; Length 1128;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 GAGGACCACCTATGT 119  
|||||  
DB 902 GAGGACCACCTATGT 916

## RESULT 10

US-09-106-217-1  
Sequence 1, Application US/09106217

Patent No. 6063576  
GENERAL INFORMATION:  
APPLICANT: Keating, Mark T.  
APPLICANT: Olson, Timothy M.  
TITLE OF INVENTION: Actin Mutations in Dilated  
TITLE OF INVENTION: Cardiomyopathy, a Heritable Form of Heart Failure  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rothwell, F19g, Ernst & Kurz, P.C.  
STREET: 555 Thirteenth Street, N.W., Suite 701 East  
CITY: Washington  
STATE: DC  
COUNTRY: U.S.A.  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/106,217  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Saxe, Stephen A.  
REGISTRATION NUMBER: 38,609  
REFERENCE/DOCKET NUMBER: 2323-125  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-783-6040  
TELEFAX: 202-783-6031  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1134 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO

ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1131  
US-09-106-217-1

Query Match 4.5%; Score 15; DB 3; Length 1134;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 GAGGACCACCTATGT 119  
|||||  
DB 908 GAGGACCACCTATGT 922

## RESULT 11

US-08-462-351-1/C  
Sequence 1, Application US/08462351

Patent No. 6106840  
GENERAL INFORMATION:  
APPLICANT: Sharma, Somesh D.  
APPLICANT: Clark, Brian R.  
APPLICANT: Lerch, Bernard L.  
TITLE OF INVENTION: MHC Conjugates Useful in Ameliorating  
TITLE OF INVENTION: Autoimmunity  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/462,351  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/210,594  
FILING DATE: 23-JUN-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/576,084  
FILING DATE: 30-AUG-1990  
APPLICATION NUMBER: US 07/690,840  
FILING DATE: 23-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/869,293  
FILING DATE: 14-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Parent, Annette S.  
REGISTRATION NUMBER: 42,058  
REFERENCE/DOCKET NUMBER: 014058-000242US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1350 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: mRNA  
FEATURE:  
NAME/KEY: CDS

LOCATION: 1.1350  
OTHER INFORMATION: /Product="acetylcholine receptor"  
OTHER INFORMATION: alpha subunit"  
US-08-462-351-1

Query Match 4.5%; Score 15; DB 3; Length 1350;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 AGACAGCAAAACGGA 63  
|||||  
DB 792 AGACAGCAAAACGGA 778

RESULT 12  
5468481-2/c  
PATENT NO. 5468481  
APPLICANT: SHARMA, SOMESH D.; CLARK, BRIAN R.; LERCH, BERNARD L.  
TITLE OF INVENTION: MHC CLASS II-PEPTIDE CONJUGATES USEFUL  
IN AMELIORATING AUTOIMMUNITY  
NUMBER OF SEQUENCES: 7  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/869,293  
FILING DATE: 14-APR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 690,840  
FILING DATE: 23-APR-1991  
APPLICATION NUMBER: 576,084  
FILING DATE: 30-AUG-1990  
APPLICATION NUMBER: 210,594  
FILING DATE: 23-JUN-1988  
APPLICATION NUMBER: 635,840  
FILING DATE: 28-DEC-1998  
APPLICATION NUMBER: 367,751  
FILING DATE: 21-JUN-1989  
SEQ ID NO: 2  
LENGTH: 1350  
5468481-2

Query Match 4.5%; Score 15; DB 6; Length 1350;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 AGACAGCAAAACGGA 63  
|||||  
DB 792 AGACAGCAAAACGGA 778

RESULT 13  
5194425-2/c  
PATENT NO. 5194425  
APPLICANT: SHARMA, SOMESH D.; LERCH, L. BERNARD; CLARK, BRIAN R.  
TITLE OF INVENTION: MHC-MEDIATED TOXIC CONJUGATES USEFUL IN  
AMELIORATING AUTOIMMUNITY  
NUMBER OF SEQUENCES: 9  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/367,751  
FILING DATE: 21-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 210,594  
FILING DATE: 23-JUN-1988  
SEQ ID NO: 2  
LENGTH: 1356  
5194425-2

Query Match 4.5%; Score 15; DB 6; Length 1356;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 AGACAGCAAAACGGA 63  
|||||  
DB 797 AGACAGCAAAACGGA 783

RESULT 14  
US-08-812-203-4  
Sequence 4, Application US/08812203  
Patent No. 5958723  
GENERAL INFORMATION:  
APPLICANT: ABRAMOVITZ, MARK  
APPLICANT: BOIE, YVES  
APPLICANT: SAWYER, NICOLE  
APPLICANT: METTERS, KATHLEEN  
APPLICANT: SLIPETZ, DEBORAH  
TITLE OF INVENTION: DNA ENCODING PROSTAGLANDIN RECEPTOR DP  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: JOHN W. WALLEN, III  
STREET: 126 E. LINCOLN AVE., P.O. BOX 2000  
CITY: RAHWAY  
STATE: NJ  
COUNTRY: USA  
ZIP: 07065-0900  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/812,203  
FILING DATE: 06-MAR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/378,682  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: WALLEN III, JOHN W.  
REGISTRATION NUMBER: 35,403  
REFERENCE/DOCKET NUMBER: MRL 94/185  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908)594-3905  
TELEX: (908)594-4/20  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1488 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-812-203-4

Query Match 4.5%; Score 15; DB 2; Length 1488;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 CTGCTGGCGCTGATG 40  
|||||  
DB 1145 CTGCTGGCGCTGATG 1159

RESULT 15  
US-09-300-864-4  
Sequence 4, Application US/09300864  
Patent No. 6214972  
GENERAL INFORMATION:  
APPLICANT: ABRAMOVITZ, MARK  
APPLICANT: BOIE, YVES  
APPLICANT: SAWYER, NICOLE  
APPLICANT: METTERS, KATHLEEN  
APPLICANT: SLIPETZ, DEBORAH  
TITLE OF INVENTION: DNA ENCODING PROSTAGLANDIN RECEPTOR DP

NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: JOHN W. WALLEN, III  
STREET: 126 E. LINCOLN AVE., P.O. BOX 2000  
CITY: RAHWAY  
STATE: NJ  
COUNTRY: USA  
ZIP: 07065-0900  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/300,864  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/378,682  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: WALLEN III, JOHN W.  
REGISTRATION NUMBER: 35,403  
REFERENCE/DOCKET NUMBER: MRL 94/185  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908)594-3905  
TELEX: (908)594-4720  
INFORMATION FOR SEQ. ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1488 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-300-864-4

Query Match 4.5%; Score 15; DB 4; Length 1488;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 CTGCTGGCGCTGATG 40  
          |||||  
DB 1145 CTGCTGGCGCTGATG 1159

Search completed: August 24, 2001, 04:34:33  
Job time: 2574 sec

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 24, 2001, 03:31:09 ; Search time 1172.14 Seconds  
(without alignments)  
2709.709 Million cell updates/sec

Title: US-08-224-621-74

Perfect score: 336  
Sequence: 1 CGTCGACATCTTTTCTATGC.....CCCCNTTGGTCCCAACCA 336

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 0

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est33:\*  
26: gb\_est34:\*  
27: gb\_est35:\*  
28: gb\_est36:\*  
29: gb\_est37:\*  
30: gb\_est38:\*  
31: gb\_est39:\*  
32: gb\_est40:\*  
33: gb\_estba:\*  
34: gb\_estfun:\*  
35: gb\_esthum1:\*  
36: gb\_esthum2:\*  
37: gb\_esthum3:\*  
38: gb\_esthum4:\*  
39: gb\_esthum5:\*  
40: gb\_esthum6:\*  
41: gb\_esthum7:\*  
42: gb\_esthum8:\*  
43: gb\_esthum9:\*

44: gb\_esthum10:\*  
45: gb\_esthum11:\*  
46: gb\_esthum12:\*  
47: gb\_esthum13:\*  
48: gb\_esthum14:\*  
49: gb\_esthum15:\*  
50: gb\_esthum16:\*  
51: gb\_esthum17:\*  
52: gb\_esthum18:\*  
53: gb\_esthum19:\*  
54: gb\_esthum20:\*  
55: gb\_esthum21:\*  
56: gb\_esthum22:\*  
57: gb\_esthum23:\*  
58: gb\_esthum24:\*  
59: gb\_esthum25:\*  
60: gb\_esthum26:\*  
61: gb\_esthum27:\*  
62: gb\_esthum28:\*  
63: gb\_esthum29:\*  
64: gb\_esthum30:\*  
65: gb\_esthum31:\*  
66: gb\_esthum32:\*  
67: gb\_esthum33:\*  
68: gb\_esthum34:\*  
69: gb\_esthum35:\*  
70: gb\_esthum36:\*  
71: gb\_esthum37:\*  
72: gb\_esthum38:\*  
73: gb\_esthum39:\*  
74: gb\_esthum40:\*  
75: gb\_esthum41:\*  
76: gb\_esthum42:\*  
77: gb\_esthum43:\*  
78: gb\_esthum44:\*  
79: gb\_esthum45:\*  
80: gb\_esthum46:\*  
81: gb\_esthum47:\*  
82: gb\_esthum48:\*  
83: gb\_esthum49:\*  
84: gb\_esthum50:\*  
85: gb\_esthum51:\*  
86: gb\_esthum52:\*  
87: gb\_esthum53:\*  
88: gb\_esthum54:\*  
89: gb\_esthum55:\*  
90: gb\_esthum56:\*  
91: gb\_esthum57:\*  
92: gb\_esthum58:\*  
93: gb\_esthum59:\*  
94: gb\_esthum60:\*  
95: gb\_esthum61:\*  
96: gb\_esthum62:\*  
97: gb\_esthum63:\*  
98: gb\_esthum64:\*  
99: gb\_esthum65:\*  
100: gb\_esthum66:\*  
101: gb\_esthum67:\*  
102: gb\_esthum68:\*  
103: gb\_esthum69:\*  
104: gb\_esthum70:\*  
105: gb\_esthum71:\*  
106: gb\_esthum72:\*  
107: gb\_esthum73:\*  
108: gb\_esthum74:\*  
109: gb\_esthum75:\*  
110: gb\_esthum76:\*  
111: gb\_esthum77:\*  
112: gb\_esthum78:\*  
113: gb\_esthum79:\*  
114: gb\_esthum80:\*  
115: gb\_esthum81:\*  
116: gb\_esthum82:\*

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

```

```

190: gb_est110:*
191: gb_est111:*
192: gb_est112:*
193: gb_est113:*
194: gb_est114:*
195: gb_est115:*
196: gb_est116:*
197: gb_est117:*
198: gb_est118:*
199: gb_est119:*
200: gb_est120:*
201: gb_est121:*
202: gb_est122:*
203: gb_est123:*
204: gb_est124:*
205: gb_est125:*
206: gb_est126:*
207: gb_est127:*
208: gb_est128:*
209: gb_est129:*
210: gb_est130:*
211: gb_est131:*
212: gb_est132:*
213: gb_est133:*
214: gb_est134:*
215: gb_est135:*
216: gb_est136:*
217: gb_est137:*
218: gb_est138:*
219: gb_est139:*
220: gb_est140:*
221: gb_est141:*
222: gb_est142:*
223: gb_est143:*
224: gb_est144:*
225: gb_est145:*
226: gb_est146:*
227: gb_est147:*
228: gb_est148:*
229: gb_est149:*
230: gb_est150:*
231: gb_est151:*
232: gb_est152:*
233: gb_est153:*
234: gb_est154:*
235: gb_est155:*
236: gb_est156:*
237: gb_est157:*
238: gb_est158:*
239: gb_est159:*
240: gb_est160:*
241: gb_est161:*
242: gb_est162:*
243: gb_est163:*
244: gb_est164:*
245: gb_est165:*
246: gb_est166:*
247: gb_est167:*
248: gb_est168:*
249: gb_est169:*
250: gb_est170:*
251: gb_est171:*
252: gb_est172:*
253: gb_est173:*
254: gb_est174:*
255: gb_est175:*
256: gb_est176:*
257: gb_est177:*
258: gb_est178:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution



## ORIGIN

Query Match 7.7%; Score 26; DB 32; Length 539;  
 Best Local Similarity 100.0%; Pred. No. 0.00033;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TTTTCTATGCTCTCCCTGCTGGCGCT 36  
 DB 393 TTTTCTATGCTCTCCCTGCTGGCGCT 418

RESULT 3  
 AV663814 539 bp mRNA EST 25-AUG-2000

LOCUS AV663814 Bos taurus brain fetus Bos taurus cDNA clone E1BR038607  
 DEFINITION 5', mRNA sequence.

ACCESSION AV663814 GI:9922844

KEYWORDS EST.

SOURCE COW.

ORGANISM Bos taurus Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Eukaryota; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovidae; Bovinae; Bos.  
 1 (bases 1 to 539)  
 Sugimoto, Y., Hirotsune, S., Takasuga, A., Itoh, R., Jitohzono, A. and  
 Suzuki, H.

REFERENCE  
 AUTHORS  
 TITLE Unpublished (2000)  
 JOURNAL Contact: Yoshikazu Sugimoto  
 COMMENT Animal Genetics Division  
 Shitakawa Institute of Animal Genetics  
 Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan  
 Tel: 81-248-25-5641  
 Fax: 81-248-25-5725  
 Email: kazusugi@cocoa.ocn.ne.jp  
 Single pass sequencing.  
 This clone was obtained from a polyA-deleted cDNA library.

## FEATURES

1..539  
 Location/Qualifiers

/organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /clone\_id="E1BR038607"  
 /clone\_lib="Bos taurus brain fetus"  
 /tissue\_type="brain"  
 /dev\_stage="fetus"  
 /lab\_host="DH10B"  
 /note="Vector: pZLI; site\_1: SalI; site\_2: NotI; Poly A  
 was deleted from a NotI site"

BASE COUNT 138 a 144 c 129 g 128 t

## ORIGIN

Query Match 7.7%; Score 26; DB 32; Length 539;  
 Best Local Similarity 100.0%; Pred. No. 0.00033;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TTTTCTATGCTCTCCCTGCTGGCGCT 36  
 DB 319 TTTTCTATGCTCTCCCTGCTGGCGCT 344

RESULT 4  
 AV663492 409 bp mRNA EST 25-AUG-2000

LOCUS AV663492 Bos taurus brain fetus Bos taurus cDNA clone E1BR036H10  
 DEFINITION 3', mRNA sequence.

ACCESSION AV663492 GI:9922522

KEYWORDS EST.

SOURCE COW.

ORGANISM Bos taurus

REFERENCE  
 AUTHORS  
 TITLE Unpublished (2000)  
 JOURNAL Contact: Yoshikazu Sugimoto  
 COMMENT Animal Genetics Division  
 Shitakawa Institute of Animal Genetics  
 Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan  
 Tel: 81-248-25-5641  
 Fax: 81-248-25-5725  
 Email: kazusugi@cocoa.ocn.ne.jp  
 Single pass sequencing.  
 This clone was obtained from a polyA-deleted cDNA library.

FEATURES  
 SOURCE 1..409  
 Location/Qualifiers

/organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /clone\_id="E1BR036H10"  
 /clone\_lib="Bos taurus brain fetus"  
 /tissue\_type="brain"  
 /dev\_stage="fetus"  
 /lab\_host="DH10B"  
 /note="Vector: pZLI; site\_1: SalI; site\_2: NotI; Poly A  
 was deleted from a NotI site"

BASE COUNT 87 a 110 c 132 g 80 t

## ORIGIN

Query Match 6.8%; Score 23; DB 32; Length 409;  
 Best Local Similarity 100.0%; Pred. No. 0.019;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 CTGCATGAGGACACACATATGTTG 121  
 DB 367 CTGCATGAGGACACACATATGTTG 345

RESULT 5  
 AV663813 414 bp mRNA EST 25-AUG-2000

LOCUS AV663813 Bos taurus brain fetus Bos taurus cDNA clone E1BR038607  
 DEFINITION 3', mRNA sequence.

ACCESSION AV663813 GI:9922843

KEYWORDS EST.

SOURCE COW.

ORGANISM Bos taurus Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Eukaryota; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovidae; Bovinae; Bos.  
 1 (bases 1 to 414)  
 Sugimoto, Y., Hirotsune, S., Takasuga, A., Itoh, R., Jitohzono, A. and  
 Suzuki, H.

REFERENCE  
 AUTHORS  
 TITLE Unpublished (2000)  
 JOURNAL Contact: Yoshikazu Sugimoto  
 COMMENT Animal Genetics Division  
 Shitakawa Institute of Animal Genetics  
 Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan  
 Tel: 81-248-25-5641  
 Fax: 81-248-25-5725  
 Email: kazusugi@cocoa.ocn.ne.jp  
 Single pass sequencing.  
 This clone was obtained from a polyA-deleted cDNA library.

FEATURES  
 SOURCE 1..414  
 Location/Qualifiers  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /clone\_id="E1BR038607"



/clone\_1lb="Bos taurus brain fetus"  
/tissue\_type="brain"  
/dev\_stage="fetus"  
/lab\_host="DH10B"  
/note="Vector: pZL1. Site\_1: SalI; Site\_2: NotI; Poly A  
was deleted from a NotI site"

BASE COUNT 87 a 112 c 133 g 82 t

ORIGIN

Query Match 6.8%; Score 23; DB 32; Length 414;  
Best Local Similarity 100.0%; Pred. No. 0.019;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 CTCGATGAGCACCACATGTGTG 121  
|||||  
Db 368 CTCGATGAGCACCACATGTGTG 346

RESULT 6  
LOCUS BE121364 435 bp mRNA EST 13-JUN-2000  
DEFINITION UI-R-CAO-bax-a-08-0-UI.s1 UI-R-CAO Rattus norvegicus cDNA clone  
UI-R-CAO-bax-a-08-0-UI 3', mRNA sequence.  
ACCESSION BE121364  
VERSION BE121364.1 GI:8513469  
KEYWORDS EST.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 435)  
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
COMMENT Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu  
Oligo-dt track not found. Not 1 site shown in beginning of sequence  
is likely internal to the message. cDNA library Preparation: M.B.  
Soares Lab Clone distribution: clones will be available through  
Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA-No.

FEATURES  
source

1. 435 location/Qualifiers  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-CAO-bax-a-08-0-UI"  
/clone\_1lb="UI-R-CAO"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker. Site\_1: Not I; Site\_2: Eco RI; The UI-R-CAO  
library is a subtracted library derived from the following  
tissues: thalamus, cerebellum, hypothalamus, medulla, pons  
, midbrain, cerebral cortex, corpus striatum, testis, and  
hippocampus. For a detailed description of the library  
from which this clone was derived, please visit our web  
site at ratest.eng.uiowa.edu. The subtraction has been  
previously described in (Bonaldo, Lennon and Soares,  
Genome Research 6:791-806, 1996)  
TAG-Seq=None found"

BASE COUNT 100 a 106 c 114 g 115 t

ORIGIN

Query Match 6.8%; Score 23; DB 163; Length 435;  
Best Local Similarity 100.0%; Pred. No. 0.019;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 ATGGAGATACAGACAGCAAAAC 60  
|||||  
Db 130 ATGGAGATACAGACAGCAAAAC 108

RESULT 7  
LOCUS BF400681 437 bp mRNA EST 28-NOV-2000  
DEFINITION UI-R-CAO-bhe-c-02-0-UI.s1 UI-R-CAO Rattus norvegicus cDNA clone  
UI-R-CAO-bhe-c-02-0-UI 3', mRNA sequence.  
ACCESSION BF400681  
VERSION BF400681.1 GI:11388656  
KEYWORDS EST.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 437)  
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
COMMENT Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu  
Oligo-dt track not found. Not 1 site shown in beginning of sequence  
is likely internal to the message. cDNA library Preparation: M.B.  
Soares Lab Clone distribution: clones will be available through  
Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA-No.

FEATURES  
source

1. 437 location/Qualifiers  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-CAO-bhe-c-02-0-UI"  
/clone\_1lb="UI-R-CAO"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker. Site\_1: Not I; Site\_2: Eco RI; The UI-R-CAO  
library is a subtracted library derived from the following  
tissues: thalamus, cerebellum, hypothalamus, medulla, pons  
, midbrain, cerebral cortex, corpus striatum, testis, and  
hippocampus. For a detailed description of the library  
from which this clone was derived, please visit our web  
site at ratest.eng.uiowa.edu. The subtraction has been  
previously described in (Bonaldo, Lennon and Soares,  
Genome Research 6:791-806, 1996)  
TAG-Seq=None found"

BASE COUNT 101 a 106 c 114 g 116 t

ORIGIN

Query Match 6.8%; Score 23; DB 148; Length 437;  
Best Local Similarity 100.0%; Pred. No. 0.019;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 ATGGAGATACAGACAGCAAAAC 60  
|||||  
Db 132 ATGGAGATACAGACAGCAAAAC 110

```

RESULT      8
LOCUS       BF394149/c
DEFINITION  BF394149 443 bp mRNA EST 27-NOV-2000
            UI-R-CAO-bha-c-12-0-UI s1 UI-R-CAO Rattus norvegicus cDNA clone
ACCESSION   BF394149
VERSION     BF394149.1 GI:11379013
KEYWORDS    EST.
SOURCE      Norway rat.
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 443)
AUTHORS    Bonaldi,M.F., Lennon,G. and Soares,M.B.
TITLE       Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Oligo-dt track not found. Not 1 site shown in beginning of sequence
            is likely internal to the message. cDNA library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com)
            Seq primer: M13 forward
            POLYA-No.

FEATURES
    source
        1..443
            location/Qualifiers
                /organism="Rattus norvegicus"
                /strain="Sprague-Dawley"
                /db_xref="taxon:10116"
                /clone="UI-R-CAO-bha-c-12-0-UI"
                /clone_1lb="UI-R-CAO"
                /lab_host="DH10B (Life Technologies)"
                /note="Vector: pTZ19-Pac (Pharmacia) with a modified
                polylinker. Site.1: Not 1; Site.2: Eco RI; The UI-R-CAO
                library is a subtracted library derived from the following
                tissues: thalamus, cerebellum, hypothalamus, medulla, pons
                and hippocampus. For a detailed description of the library
                from which this clone was derived, please visit our web
                site at ratest.eng.uiowa.edu. The subtraction has been
                previously described in (Bonaldi, Lennon and Soares,
                Genome Research 6:791-806, 1996)
                TAG-Seq=None found"
BASE COUNT  102 a 106 c 117 g 118 t
ORIGIN
Query Match 6.8%; Score 23; DB 148; Length 443;
Best Local Similarity 100.0%; Pred. No. 0.019;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 38 ATGGAGATACAGACGACAAAC 60
|||||
Db 138 ATGGAGATACAGACGACAAAC 116

RESULT      9
LOCUS       BE032856 547 bp mRNA EST 09-JUL-2000
DEFINITION  BE032856 MARC 1Pig Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BE032856
VERSION     BE032856.1 GI:8327865
KEYWORDS    EST.
SOURCE      pig.

```

```

ORGANISM    Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
REFERENCE   1 (bases 1 to 547)
AUTHORS    Fahrnenkrug,S.C., Fekking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.W., Bennett,G.A., Laegreid,W.W.
            and Keefe,J.W.
TITLE       Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
JOURNAL     Unpublished (2000)
COMMENT     Contact: Smith TP
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@mail.marc.usda.gov
            Single pass sequencing. Bases called and alt. trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -mismatch 12 options.
            PCR primers
            FORWARD: AGGAACAGCATGACCAT
            BACKWARD: GTTTCCTCAGTCACGACG
            Plate: 66 row: K column: 23
            Seq primer: ATTGAGTGACACTATAG.
            Location/Qualifiers
                1..547
                /organism="Sus scrofa"
                /db_xref="taxon:9823"
                /clone_1lb="MARC 1Pig"
                /tissue-type="pooled"
                /lab_host="DH10B"
                /note="Vector: pCMV SPORT6; site.1: XbaI; site.2: XhoI;
                library made from pooled tissue from day 11, 13, 15, 20,
                and 30 embryos."
FEATURES
    source
        1..547
            location/Qualifiers
                /organism="Sus scrofa"
                /db_xref="taxon:9823"
                /clone_1lb="MARC 1Pig"
                /tissue-type="pooled"
                /lab_host="DH10B"
                /note="Vector: pCMV SPORT6; site.1: XbaI; site.2: XhoI;
                library made from pooled tissue from day 11, 13, 15, 20,
                and 30 embryos."
BASE COUNT  110 a 178 c 149 g 110 t
ORIGIN
Query Match 6.8%; Score 23; DB 162; Length 547;
Best Local Similarity 100.0%; Pred. No. 0.019;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 11 TTTTCATGCTCTCCCTGCTGCG 33
|||||
Db 6 TTTTCATGCTCTCCCTGCTGCG 28

```

```

RESULT      10
LOCUS       BG293757 711 bp mRNA EST 21-FEB-2001
DEFINITION  BG293757 602390647P1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4502603 5',
            mRNA sequence.
ACCESSION   BG293757
VERSION     BG293757.1 GI:13053739
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 711)
AUTHORS    NIH-MGC http://mgc.ncl.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgaabs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
            cDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10371 row: n column: 12

```



Email: [segreff@genoscope.cns.fr](mailto:segreff@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Location/Qualifiers  
1  
OC

For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)

```

FEATURES
source
Location/Qualifiers
1. .179
/organism="Homo sapiens"

```

```

/clone="CS0DF0Z8Y0L8"
/clone.lib="LRL FL013_Fbn1"
/dev_stage="pooled tissue from post conception fetuses (20
week, 24 week and 26 week)"
lab_host="DH10B"

```

1A10, 10B5 - DNA  
/note="Organ: Fetal brain, Vector: pCMVSPORT 6; 1st strand  
cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-stranded cDNA was digested with Not I  
and cloned into the Not I and Eco RV sites of the  
pCMVSPORT 6 vector. Library was constructed by Life  
Technologies. Contact : Feng Liang Life Technologies, a  
division of Invitrogen 9800 Medical Center Drive Rockville  
Maryland 20850, USA Fax : (1) 301 610 8371 Email :  
filiang@lifetech.com URL :  
http://fulllength.invitrogen.com"

BASE COUNT	229 a	251 c	224 g	245 t	3 others
ORIGIN					

Query Match	6.5%;	Score 22;	DB 106;	Length 952;
Best Local Similarity	100.08;	Pred. No. 0.079;	0.079;	Gaps 0

```

QY      11  TTTTCTATGCTCTCCCTGCTGG  32
          |||||
Db      299  TTTTCTATGCTCTCCCTGCTGG  32

```

[illegible]

REFERENCE  
AUTHORS

Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Mollusca; Lophotrochozoa; Cephalopoda; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 179)

Adams, M.D., Kellavage, A.R., Fleischmann, R.D., Fulmer, K.A., Butt, C.J., Lee, N.H., Kirkness, E.F., Weissmann, R.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Well, C., Clayton, R.A., Cline, T.R., Cotton, M.D., Earle, H.J., Geoghegan, N.S., Glodde, A., L.M., Fitzhugh, W.M., Fitchman, J.L., Geoghegan, N.S., Glodde, A., Kellum, J.C., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmorek, S.M., Merrick, J.M., Pelligrino, S.M., Moreno-Palancques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Uterback, T.R., Weidman, J.F., Li, X., Bedarkik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dimke, D., Feng, D.-F., Ferrite, A., Fischer, C., Hastings, G.A., He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hungenjurg, J., Li, H., Meisner, P.S., Olsen, H., Raymond, J., Wel, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannone, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M., and Venter, J.C.

**TITLE** Initial assessment of human gene diversity and conservation based upon 83 million nucleotides of cDNA sequences  
**JOURNAL** Nature 377 (6547 Suppl.), 3-174 (1995)

JOURNAL  
MEDLINE  
COMMENT

The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: [arkerlavetlgr.org](mailto:arkerlavetlgr.org)

BASE COUNT	23 a	52 c	58 g	44 c	2000
ORIGIN					

Query Match	6.08;	Score 20;	DB 5;	Length 179;
Best Local Similarity	100.08;	Pred. No. 1;		
Matches	20;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0;

QY	22	CTCCCTGCTGGCGCTGATGC	41
Db	148	CTCCCTGCTGGCGCTGATGC	167

[illegible]

ORGANISM	REFERENCE	AUTHORS
Homo sapiens	1 (bases 1 to 371)	Miller, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., et al.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		

TITLE	WashU-NCI human ESI Project
JOURNAL	Unpublished (1997)
COMMENT	Contact: Wilson RK Washington University School of Medicine

Contact: Watson, Paul  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)  
This clone is available royalty-free through LINT; contact the  
IMaGE consortium ([info@image.jhmi.gov](mailto:info@image.jhmi.gov)) for further information.  
Insert Length: 1246 Std Error: 0.00  
Seq primer: 28m13 rev2.Er from Amersham  
High quality sequence stop: 329

FEATURES  
source

1. .371  
/-----"Homo sapiens"

```

/organism="Homo sapiens"
/db_xref="db:5939614"
/db_xref="taxon:9606"
/clone="IMAGE:739933"
/clone_lib="Scars ovary tumor NBDH9"
/sex="Female"
/tissue_type="ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
/notice="Organ: Ovary; Vector: pT73D (Pharmacia) 1st modified polylinker; Site_1: Not I; Site_2: Eco RI; strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCATCTGTAAGAGGCGGCGCGCGGTTTTTTTTTTT 3']"
/sequence="cDNA was size selected, ligated to Eco RI double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into

```

the Not I and Eco RI sites of a modified pT7T3 vector  
(Pharmacia). Library constructed by Bento Soares and  
M. Fatima Bonaldo."

BASE COUNT 65 a 136 c 89 g 81 t  
ORIGIN

Query Match 6.0%; Score 20; DB 8; Length 371;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 CTCCTCTGCTGGCGCTGATGG 41  
|||||  
Db 24 CTCCTCTGCTGGCGCTGATGG 43

Search completed: August 24, 2001, 04:13:18  
Job Time: 2529 sec

**This Page Blank (uspto)**  
**Best Available Copy**

**This Page Blank (uspto)**